UNIVERSIDAD AUTONOMA DE MADRID

Facultad de Ciencias
Departamento de Biología Molecular

Structural basis for Cbl-induced SH3 oligomerization

Memoria presentada por Nayra Cárdenes Estellés para optar al grado de Doctora en Ciencias.

Este trabajo fue realizado en el Centro Nacional de Investigaciones Oncológicas (CNIO) bajo la dirección del Dr. Jerónimo Bravo Sicilia.

Financiado por la Fundación Ramón Areces.

Centro Nacional de Investigaciones Oncológicas

UNIVERSIDAD AUTONOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular

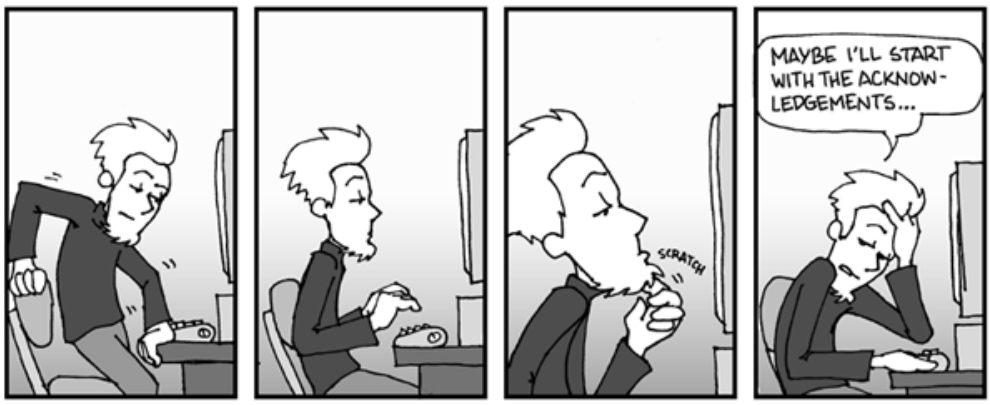
El Dr. Jerónimo Bravo Sicilia, doctor en Ciencias Biológicas-Bioquímica por la Universitat Politécnica de Catalunya

CERTIFICA:

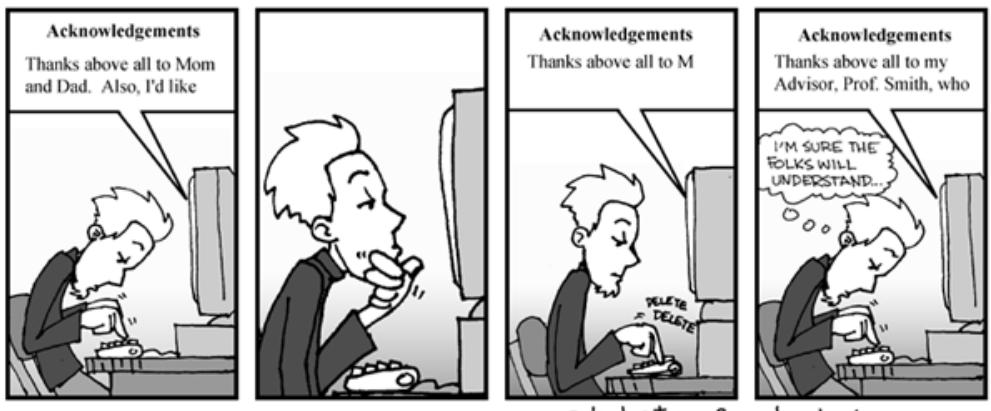
Que la tesis doctoral titulada “Structural basis for Cbl-induced SH3 oligomerization”, ha sido realizada en el Centro Nacional de Investigaciones Oncológicas (CNIO) y tutelada en el Departamento de Biología Molecular de la Universidad Autónoma de Madrid.

Como director de esta tesis, realizada por Nayra Cárdenes Estellés, está de acuerdo con los métodos empleados y con los resultados obtenidos. Por tanto considera que reúne los requisitos de originalidad y calidad necesarios para ser presentada y defendida ante un tribunal con el fin de optar al grado de Doctora.

Fdo.: Dr. Jerónimo Bravo Sicilia



A Álvar,
quien despertó en mí
el interés por la ciencia


Acknowledgements
Thanks above all to Mom and Dad. Also, I'd like
Acknowledgements
Thanks above all to M
DELETE
DELETE
Acknowledgements
Thanks above all to my Advisor, Prof. Smith, who
I'M SURE THE FOLKS WILL UNDERSTAND...
phd.stanford.edu/comics

Agradecimientos/Acknowledgements

Quiero expresar mi agradecimiento a mi director de tesis Jerónimo Bravo, quien me enseñó que la biología no es matemáticas. Me siento afortunada por la oportunidad que me brindó para participar en este interesante proyecto y en especial por su apoyo al final de este largo camino.

A mis padres (quienes me dieron la oportunidad de estudiar lo que más me gusta), mi hermano Dani y la Madrina, por su apoyo incondicional y su fe ciega. Por estar en los buenos y los malos momentos y alentarme siempre a superarme.

A toda mi familia, por quererme y porque estuvieron, están y siempre estarán a mi lado de una u otra manera. A mis abuelos, a quienes les hubiera gustado acompañarme en estos momentos. A Pepe y Charo, Carmen y Justo, y a la familia Sevilla-Kreysa, quienes siempre confiaron en mí.

A todos los miembros del 007, quienes no solo me han regalado sus conocimientos prácticos y teóricos sin los cuales este trabajo no hubiera sido posible, sino que han sido como una segunda familia. A *"los mayores"* José, Carme y Miguel; a *"los menos mayores"* Rebe y Fede; y a *"las niñas"*, Mer, Lina, Elena y Carol. En especial a Gabi, mi guía en este trabajo del cual es parte esta tesis y que impulsó este proyecto, y a Miguel, por su paciencia y sus enseñanzas en cristalografía. A todos ellos, los de ahora, los de antes y los que pasaron esporádicamente dejando su huella, gracias por aguantarme hasta en mi modo "payaso loco", por estar siempre disponibles para cualquier duda y por hacer el día a día más ameno.

A mi compañera de batallas dentro y fuera del laboratorio, Mer. Yo de mayor quiero ser como tú. A Fátima y Susana, que me acompañaron en mi re-adaptación a España y me prestaron ideas para el formato de esta tesis. A todos mis amigos, los del lado de acá, Diego, Javi, Rubén, Erika, José, María, Martos, JuanPa, Carmen, Elena, María G…. y los del lado de allá, Ome, Maris, Kenks, JuKlu, Minili, Rubén, Yoko… Gracias por quererme a pesar de todo.

A todos los miembros del Programa, en especial a los "erremeneros" y los cristalógrafos, por su apoyo moral, científico y por las sonrisas regaladas en los pasillos. A Pascal, que me ayudó con los experimentos de fluorimetría; a Dani y Paco, por los experimentos de RMN y Jesús por la ayuda con el DC y el ITC. A los bioinfos de antes, que me adoptaron como una más, y a los de ahora, todos ellos siempre dispuestos a ayudar y a pasarla bien. A Ángel y Edu por los momentos divertidos. A Paula, con quien intentamos los ensayos de Biacore. A todos en general, gracias.

To Janet L. Smith, who has always put her trust on me, and has supported me even after so many years. To Sonita, who has always been close, even through the distance.

A Pedro García, quien me alentó en mis primeras inmersiones en la ciencia. A Paquito, que conseguía quitarme el estrés del día a día. En definitiva a todos los que han participado en este proceso.

Y finalmente mi agradecimiento más cariñoso a Manuel, porque cuando decidió estar, estuvo al 200%. Porque él me enseñó que hasta los errores más difíciles se pueden corregir, por su apoyo, su cariño y su dedicación. Gracias.

INDEX

I. RESUMEN 1

II. INTRODUCTION 7

1. EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR) NEGATIVE REGULATION 8
1.1. THE FAMILY OF ADAPTOR PROTEINS CIN85/CMS 10
1.2. MOLECULAR INSIGHTS INTO THE SH3 DOMAIN-MEDIATED INTERACTIONS 12
1.3. IDENTIFICATION OF A NOVEL CIN85/CMS SH3 INTERACTING MOTIF IN CBL PROTEINS 16

III. OBJECTIVES 21

IV. MATERIALS AND METHODS 25

1. CLONING 25
1.1. DNA AMPLICON GENERATION 26
1.2. DNA DIGESTION AND DEPHOSPHORYLATION 28
1.3. DNA LIGATION 28
1.4. DNA TRANSFORMATION 29
1.4.1. PREPARATION OF TRANSFORMATION COMPETENT *E. COLI* CELLS 29
1.4.2. TRANSFORMATION OF *E. COLI* CELLS BY ELECTROPORATION OR HEAT SHOCK 30
1.5. CONFIRMATION OF THE DNA SEQUENCE 30
1.6. DNA DETECTION 31
1.6.1. DNA ELECTROPHORESIS 31
1.6.2. DNA QUANTIFICATION 31
1.7. DNA PURIFICATION 31
2. HETEROLOGOUS PROTEIN OVER-EXPRESSION AND SOLUBILIZATION 32
3. PROTEIN PURIFICATION 34
3.1. HIS-TAGGED AND GST-TAGGED PROTEIN PURIFICATION 35
3.2. UNTAGGED CIN85 AND CMS SH3A PURIFICATION 37
4. PROTEIN DETECTION 37
4.1. SDS-PAGE 37
4.2. PROTEIN QUANTIFICATION 38
5. CRYSTALLOGRAPHY 38
5.1. PROTEIN CRYSTALLIZATION 39
5.2. CRYSTAL CRYO-COOLING 42
5.3. CRYSTAL DATA COLLECTION 43
5.3.1. DATA REDUCTION 45
5.3.2. STRUCTURE SOLUTION BY MOLECULAR REPLACEMENT (MR) 45
5.3.3. MODEL REFINEMENT 47
5.4. PROTEIN CHARACTERIZATION 48

5.4.1. Fluorescence Spectroscopy 48
5.4.2. Isothermal Titration Calorimetry (ITC) 49
5.4.3. Circular Dichroism (CD) 49
5.4.4. Chemical shifts in monodimensional Nuclear Magnetic Resonance (NMR) 50

V. RESULTS 53

1. CIN85 53
1.1. Individual CIN85 SH3 domains 53
1.1.1. Design of SH3 constructs and the Cbl peptides 53
1.1.2. Purification, crystallization and structure of CIN85-SH3C$_{270-328}$His$_6$ 58
1.1.2.1. CIN85-SH3C binds to c-Cbl peptides 62
1.1.2.2. Structure of CIN85-SH3C$_{270-328}$His$_6$ 64
1.1.3. Purification, characterization, crystallization and structure of CIN85-SH3A$_{2-58}$ 70
1.1.3.1. Structure of CIN85-SH3A$_{2-58}$:Cbl-b$_{902-912}$ complex 75
V.1.1.3.1.1. Crystallization and structure of the CIN85-SH3A$_{2-58}$:CD2$_{324-333}$ complex 85
2. CMS 91
2.1. The N-terminal SH3 domain of CMS is the closest to CIN85-SH3A 92
2.1.1. CMS-SH3A:Cbl-b complex 92
2.1.1.1. CMS-SH3A:Cbl-b$_{902-912}$ structrure refinement 92
2.1.1.2. CMS-SH3A:YCbl-b$_{903-914}$ structure 96
2.1.1.3. CMS-SH3A binding to different length Cbl-b derived peptides 99

VI. DISCUSSION 103

1. CIN85 103
1.1. CIN85 SH3 domains 103
1.2. CIN85-SH3A:Cbl-b structure 105
1.3. CIN85-SH3A:CD2 structure 107
2. CMS 108
2.1. CMS-SH3A:Cbl-b forms a heterotrimeric complex with a differential recognition from CIN85-SH3A:Cbl-b 108
3. CIN85 and CMS SH3 binding consensus motifs 112
4. Other examples of Ligand Induced SH3 Oligomerization (LISO) 114
4.1. Comparison of all LISO arrangements 116

VII.A CONCLUSIONS 121

VII.B CONCLUSIONES 122

VIII. BIBLIOGRAPHY 125

IX. APPENDIX I: SUPPLEMENTAL MATERIAL 133

X. APPENDIX II: ABBREVIATIONS 135

XI. APPENDIX III: ARTICLES 137

FIGURES

INTRODUCTION

FIGURE II.1. SCHEMATIC REPRESENTATION OF THE DOMAIN STRUCTURE OF HUMAN CBL PROTEINS. 10

FIGURE II.2. SCHEMATIC REPRESENTATION OF THE DOMAIN STRUCTURE OF CIN85 AND CMS. 11

FIGURE II.3. SCHEMATIC DIAGRAM OF SH3 FOLD. 13

FIGURE II.4. STRUCTURE AND BINDING MECHANISM OF SH3 DOMAINS. 14

FIGURE II.5. "THE STRUCTURES OF CLASS I AND CLASS II PEPTIDE LIGANDS ARE DEPICTED SCHEMATICALLY. 15

FIGURE IV.1. EXAMPLE OF INTENSITY CALCULATION FOR A VECTOR AND AN INSERT USED FOR DNA LIGATION. 28

FIGURE IV.2. DIAGRAM OF HANGING DROP METHOD. 40

FIGURE IV.3. DIAGRAM OF SITTING DROP METHOD. 40

FIGURE IV.4. REPRESENTATION OF PAIRS OF BRAGG PLANES. 44

FIGURE IV.5. THE EWALD SPHERE CONSTRUCTION. 44

FIGURE IV.6. A PARTICULARLY DRAMATIC ILLUSTRATION OF THE IMPORTANCE OF PHASE. 46

FIGURE V.1 MULTIPLE CLONING SITE (MCS) OF PET-21A FROM NOVAGEN. 54

FIGURE V.2. OVERALL STRUCTURE OF THE C-TERMINAL SH3 DOMAIN OF SEM5 BOUND TO A POLYPROLINE RICH SOS DERIVED PEPTIDE (PDB CODE 1SEM). 55

FIGURE V.3. ALIGNMENT OF THE NEWLY DESIGNED SH3 CONSTRUCT SEQUENCES FROM CIN85 AND THE C-TERMINAL SH3 OF SEM5. 55

FIGURE V.4. OVEREXPRESSION OF CIN85-SH3B$_{101-157}$HIS$_6$ IN *E. COLI* STRAINS AT 37ºC. 56

FIGURE V.5. PURE CIN85-SH3B$_{101-157}$HIS$_6$. 56

FIGURE V.6. CIRCULAR DICHROISM (CD) SPECTRA OF CIN85-SH3B$_{101-257}$HIS$_6$ AND OTHER SH3 DOMAINS. 57

FIGURE V.7. THERMAL STABILITY CURVE OF CIN85-SH3B$_{101-257}$HIS$_6$ MONITORED BY FAR-UV CD. 57

FIGURE V.8. DETAIL OF THE MONODIMESIONAL NUCLEAR MAGNETIC RESONANCE (NMR) SPECTRA OF THE CIN85-SH3B$_{101-157}$HIS$_6$ PROTEIN CONSTRUCT BETWEEN 6 AND 5.5 PPM. 57

FIGURE V.9. OVEREXPRESSION OF CIN85-SH3C$_{270-328}$HIS$_6$. 58

FIGURE V.10. PURIFICATION OF -SH3C$_{270-328}$HIS$_6$ IN A NICKEL CHARGED CHELATING HITRAP COLUMN. 59

FIGURE V.11. PURIFICATION OF CIN85-SH3C$_{270-328}$HIS$_6$ IN A Q FF HITRAP COLUMN. 60

FIGURE V.12. PURIFICATION OF CIN85-SH3C$_{270-328}$HIS$_6$ BY SIZE EXCLUSION. 60

FIGURE V.13. CIRCULAR DICHROISM (CD) SPECTRA OF CIN85-SH3C$_{270-328}$HIS$_6$. 61

FIGURE V.14. DETAIL OF THE MONODIMESIONAL NUCLEAR MAGNETIC RESONANCE (NMR) SPECTRA OF THE CIN85-SH3C$_{270-328}$HIS$_6$ PROTEIN CONSTRUCT BETWEEN 6 AND 5.5 PPM. 62

FIGURE V.15. BINDING OF C-CBL PEPTIDES TO CIN85-SH3C$_{273-326}$HIS$_6$ CONSTRUCT MONITORED BY COLLISIONAL QUENCHING OF CIN85-SH3C FLUORESCENCE. 63

FIGURE V.16. PUBLISHED FLUORESCENCE DATA OF OTHER SH3 DOMAINS BINDING TO POLYPROLINE PEPTIDES. 63

FIGURE V.17. PURIFIED CIN85-SH3C$_{270-328\ TAA\ TGA}$ (MIDDLE BAND) AND C-CBL$_{820-830}$ AFTER TAG CLEAVAGE (LOWER BAND). 64

FIGURE V.18. CIN85-SH3C$_{270-328}$HIS$_6$ PROTEIN CRYSTALS. 64

FIGURE V.19. STRUCTURE OF CIN85-SH3C$_{270-328}$HIS$_6$. 66

FIGURE V.20. RAMACHANDRAN PLOT OF CIN85-SH3CHIS$_6$ STRUCTURE VALIDATION IN MOLPROBITY WEBTOOL. 66

FIGURE V.21. DETAIL OF THE B-BULGE IN THE CIN85-SH3C$_{270-328}$HIS$_6$ STRUCTURE. 67

FIGURE V.22 DETAIL OF THE RT LOOP IN THE CIN85-SH3C$_{270-328}$HIS$_6$ STRUCTURE. 68

FIGURE V.23. SURFACE REPRESENTATION OF THE CIN85-SH3C$_{270-328}$HIS$_6$ STRUCTURE. 69

FIGURE V.24. CRYSTALS OBTAINED WITH A PROTEIN MIX OF CIN85-SH3C$_{270-328}$HIS$_6$ AND CBL-B$_{902-912}$ AT 4ºC. 69

FIGURE V.25. OVEREXPRESSION OF CIN85-SH3A$_{2-58}$HIS$_6$ IN *E. COLI* ROSETTA(DE3)PLYSS STRAIN. 70

FIGURE V.26. PURIFICATION OF CIN85-SH3A$_{2-58}$HIS$_6$ IN A NICKEL CHARGED CHELATING HITRAP COLUMN. 71

FIGURE V.27. PURIFICATION OF CIN85-SH3A$_{2-58}$HIS$_6$ IN A Q HP HITRAP COLUMN. 71

FIGURE V.28. PURE CIN85-SH3A$_{2-58}$HIS$_6$. 72

FIGURE V.29. CIRCULAR DICHROISM (CD) SPECTRA OF CIN85-SH3A$_{2-58}$HIS$_6$ (A) AND DETAIL OF THE MONODIMESIONAL NUCLEAR MAGNETIC RESONANCE (NMR) SPECTRA OF THE CIN85-SH3A$_{2-58}$HIS$_6$ PROTEIN CONSTRUCT BETWEEN 6 AND 5.5 PPM. 72

FIGURE V.30. ISOTHERMAL TITRATION CALORIMETRY MEASUREMENTS OF COMPLEX FORMATION BETWEEN CIN85-SH3A (50MM) AND THE CBL-B PEPTIDE (545MM). 73

FIGURE V.31. IMAGE OF CIN85-SH3A$_{2-58}$HIS$_6$ CRYSTAL DIFFRACTION. 74

FIGURE V.32. ALIGNMENT OF CIN85 SH3A AND SH3C SEQUENCES. 74

FIGURE V.33. OVEREXPRESSION OF CIN85-SH3A$_{2-58\ TGA}$. 75

FIGURE V.34. 4-23% POLYACRYLAMIDE GRADIENT SDS-PAGE SHOWS THE $(NH_4)_2SO_4$ PRECIPITATION OF CIN85-SH3A$_{2-58\ TGA}$. 75

FIGURE V.35. PURIFICATION OF CIN85-SH3A$_{2-58\ TGA}$ BY HYDROPHOBIC INTERACTION CHROMATOGRAPHY. 76

FIGURE V.36. PURE CIN85-SH3A$_{2-58\ TGA}$. 76

FIGURE V.37. CRYSTALS OF OF CIN85-SH3A$_{2-58\ TGA}$ OBTAINED AT 22^{o}C. 77

FIGURE V.38. CRYSTALS OF OF CIN85-SH3A$_{2-58\ TGA}$ USED FOR STRUCTURE DETERMINATION. 78

FIGURE V.39. STATISTIC ANALYSIS OF THE INTENSITIES OF TWINNED DATA OBTAINED WITH THE SFCHECK TOOL. 79

FIGURE V.40. A HYPOTHETICAL EXAMPLE OF MEROHEDRAL TWINNING IN WHICH SEPARATE DOMAINS WITH P3 SYMMETRY ARE RELATED BY A 180° ROTATION ABOUT AN AXIS PARALLEL TO THE THREEFOLD CRYSTALLOGRAPHIC AXIS. 79

FIGURE V.41. OVERALL STRUCTURE OF THE CIN85-SH3A:CBL-B HETEROTRIMERIC COMPLEX. 81

FIGURE V.42. SCHEMATIC REPRESENTATION OF THE PROTEIN-PEPTIDE CONTACTS BETWEEN THE CBL PEPTIDE AND CIN85-SH3A. 82

FIGURE V.43. DETAIL OF THE CIN85-SH3A:CBL-B HETEROTRIMERIC STRUCTURE. 83

FIGURE V.44. DETAIL OF THE HYDROPHOBIC POCKET II IN SH3 II OF THE CIN85-SH3A:CBL-B HETEROTRIMERIC STRUCTURE. 83

FIGURE V.45. COMPARISON OF THE SH3 I:CBL-B AND SH3 II:CBL-B INTERFACES FROM THE CIN85-SH3A:CBL-B HETEROTRIMER COMPLEX. 84

FIGURE V.46. RAMACHANDRAN PLOT OF CIN85-SH3A:CBL-B STRUCTURE VALIDATION IN MOLPRPB. 85

FIGURE V.47. CRYSTALS OF CIN85-SH3A$_{2\text{-}58\ TGA}$:CD2$_{324\text{-}333}$ COMPLEX. 86

FIGURE V.48. DIFFRACTION IMAGE FROM THE CRYSTAL OF THE COMPLEX OF CIN85-SH3A$_{2\text{-}58\ TGA}$:CD2$_{324\text{-}333}$. 86

FIGURE V.49. SELF ROTATION FUNCTION OF DATA FROM CIN85-SH3A:CD2 CRYSTAL. 87

FIGURE V.50. PLOTTED I/Σ VS EACH REFLECTION AXIS: (A) H, (B) K, (C) L. SYSTEMATIC ABSENCES ARE DETECTED EVERY TWO REFLECTIONS IN THE K AND L AXES. 87

FIGURE V.51. PRELIMINARY MODEL OF THE CIN85-SH3A:CD2 COMPLEX AFTER AUTOMATIC REBUILDING IN ARP/WARP. 88

FIGURE V.52. RAMACHANDRAN PLOT OF CIN85-SH3A:CD2 STRUCTURE VALIDATION IN MOLPROBITY WEBTOOL. 89

FIGURE V.53. OVERALL STRUCTURE OF THE CIN85-SH3A:CD2 HETERODIMERIC COMPLEX. 90

FIGURE V.54. CIN85-SH3A:CD2 INTERFACE. SURFACE OF CIN85-SH3A DOMAIN IS COLORED IN GREY, AND THE CD2 PEPTIDE IS SHOWN IN GREEN STICKS. 90

FIGURE V.55. BINDING OF C-CBL PEPTIDE TO CMS-SH3A AND CMS-SH3B MONITORED BY COLLISIONAL QUENCHING OF CMS SH3 FLUORESCENCE. 91

FIGURE V.56. SEQUENCE ALIGNMENT OF THE N-TERMINAL SH3 DOMAINS FROM CMS AND CIN85. 92

FIGURE V.57. RAMACHANDRAN PLOT OF CMS-SH3A:CBL-B STRUCTURE COMPLEX VALIDATION IN MOLPROB PROGRAM. 93

FIGURE V.58. SELF ROTATION FUNCTION OF DATA FROM CMS-SH3A:CBL-B CRYSTAL. 93

FIGURE V.59. STEREO VIEW OF SYMMETRICALLY RELATED PEPTIDES FROM THE CMSA:CBL-B HETEROTRIMERIC COMPLEX SHOWS THE DOUBLE ORIENTATION OF THE PEPTIDE. 94

FIGURE V.60. OVERALL STRUCTURE OF THE CMS-SH3A:CBL-B HETEROTRIMERIC COMPLEX. 94

FIGURE V.61. SCHEMATIC REPRESENTATION OF CONTACTS BETWEEN CMS-SH3A AND THE CBL-B PEPTIDE (LIGHT BLUE). 95

FIGURE V.62. CRYSTALS OF THE CMS-SH3A:YCBL-B$_{903\text{-}914}$. 96

FIGURE V.63. CRYSTAL OF CMS-SH3A:YCBL-B$_{903\text{-}914}$ COMPLEX, DIFFRACTED IN THE SYNCHROTRON RADIATION SOURCE AT THE ESRF FACILITY AT BEAMLINE ID-14. 96

FIGURE V.64. IMAGES OF DATA COLLECTION OF THE CRYSTAL OF THE CMS-SH3A:YCBL-B$_{903\text{-}914}$ COMPLEX WITH A HIGH RESOLUTION LIMIT OF 1.4Å (LEFT) AND 3.5Å (RIGHT). 97

FIGURE V.65. RAMACHANDRAN PLOT OF CMS-SH3A:YCBL-B$_{903\text{-}914}$ STRUCTURE COMPLEX VALIDATION WITH THE MOLPROBITY WEB-BASED TOOL. 98

FIGURE V.66. CMS-SH3A:YCBL-B$_{903\text{-}914}$ STRUCTURE. 99

FIGURE VI.1. SEQUENCE ALIGNMENT OF ALL CIN85 AND CMS SH3 DOMAINS. 104

FIGURE VI.2. STRUCTURAL ALIGNMENT OF CIN85-SH3A AND CIN85-SH3C DOMAINS. 104

FIGURE VI.3. ALIGNMENT OF INTERACTING SEQUENCES FROM CIN85 NATURAL PARTNERS. 106

FIGURE VI.4. COMPARISON OF THE CIN85-SH3A:CBL-B AND CMS-SH3:CBL-B TERNARY COMPLEXES. 109

FIGURE VI.5. COMPARISON OF THE SH3 I:CBL-B (CLASS I-LIKE) INTERFACE FROM CIN85-SH3A AND CMS-SH3A. 110

FIGURE VI.6. CBL-B RESIDUES IMPLICATED IN CLASS I-LIKE AND CLASS II-LIKE RECOGNITION BY SH3A FROM CIN85 AND CMS. 113

FIGURE VI.7. COMPARISON OF SH3:POLYPROLINE HETEROTRIMERS. 117

FIGURE VI.8. SEQUENCE COMPARISON OF SH3:POLYPROLINE HETEROTRIMER COMPONENTS. 118

TABLES

MATERIALS AND METHODS

TABLE IV.1. ENZYMES, BUFFERS AND REAGENTS USED FOR DNA CLONING. 26
TABLE IV.2. OLIGONUCLEOTIDES USED FOR THE AMPLICON GENERATION OF THE CIN85 SH3 CONSTRUCTS 27
TABLE IV.3. *E. COLI* EXPRESSION VECTORS. 29
TABLE IV.4. *E. COLI* STRAINS 32
TABLE IV.5. SOLUBILIZATION BUFFERS USED FOR CELL LYSIS. 34
TABLE IV.6. PURIFICATION SCHEMES FOR CIN85-SH3 CONSTRUCTS AND CMS-SH3A. 36
TABLE IV.7. CRYSTALLIZATION SCREENS USED FOR INICIAL SEARCH OF CRYSTALLIZATION CONDITIONS. 40
TABLE IV.8. CRYSTALLIZATION PLATES USED FOR CRYSTAL GROWTH. 41
TABLE IV.9. POLYPROLINE PEPTIDES USED FOR CO-CRYSTALIZATION AND BINDING STUDIES. 42

RESULTS

TABLE V.1. CRYSTALLOGRAPHIC DATA COLLECTION AND REFINEMENT ANALYSIS OF CIN85-SH3CHIS STRUCTURE 65
TABLE V.2. CRYSTALLOGRAPHIC DATA COLLECTION AND REFINEMENT STATISTICS OF CIN85-SH3A$_{2\text{-}58\ TGA}$:CBL-B$_{902\text{-}912}$ STRUCTURE 80
TABLE V.3. CRYSTALLOGRAPHIC DATA COLLECTION AND ANALYSIS OF CIN85-SH3A$_{2\text{-}58\ TGA}$:CD2$_{324\text{-}333}$ STRUCTURE 89
TABLE V.4. CRYSTALLOGRAPHIC DATA COLLECTION AND ANALYSIS 93
TABLE V.5. CRYSTALLOGRAPHIC DATA COLLECTION AND ANALYSIS OF CMS-SH3A:YCBL-B$_{903\text{-}914}$ STRUCTURE. 98

RESUMEN

I. Resumen

La célula, como estructura anatómica y funcional fundamental de la materia viva, necesita llevar a cabo funciones esenciales para la supervivencia y el desarrollo. Cualquier ambigüedad o tendencia a alejarse de su objetivo produce graves efectos tanto en la célula como en el organismo del que forma parte. Por esta razón, para el correcto funcionamiento y desarrollo a nivel uni y pluricelular, se requiere de un control riguroso de la morfología, el crecimiento y la homeostasis celulares. Todo ello depende de la estricta regulación de las respuestas a estímulos extracelulares que modulan el nivel de activación celular.

La unión de factores o ligandos extracelulares a sus receptores provocan un amplio repertorio de complejos multiproteicos altamente dinámicos. De esta forma, la señal se amplifica y se traduce en cambios en el proteoma y el genoma celulares.

La mayoría de receptores de membrana producen señales positivas, y la alteración de la intensidad y de la duración de estas señales provoca cambios en las funciones de la célula. Para asegurar que se alcanzan y mantienen unos niveles adecuados de señalización, las células han desarrollado mecanismos específicos de señalización negativa.

Con el objetivo de atenuar rápidamente la transmisión de señales a la célula, se produce una serie compleja de eventos en cadena. La naturaleza ha desarrollado un mecanismo efectivo para asegurar de manera rápida y eficaz la terminación de señales mediante la insensibilización a estímulos externos. Para ello, se lleva a cabo la internalización de los receptores de membrana encargados de transmitir las señales del exterior al interior de la célula.

El factor de crecimiento epidérmico EGF (del inglés: *Epidermal Growth Factor*) y su receptor EGFR han sido identificados como componentes clave del proceso de crecimiento y replicación celulares. El aumento en actividad del receptor EGFR se ha visto directamente implicado en una serie de tumores sólidos como el cáncer de colon, de mama, gástrico y ovárico. La unión de EGF al receptor EGFR da lugar a su activación inicial y consecuente supresión tras una rápida eliminación de la superficie celular mediante vesículas envueltas de clatrina. El estudio de este proceso endocítico que modula la transducción de señales vía EGFR podría ofrece nuevas oportunidades terapéuticas y elucidar mecanismos que contribuyan a la eficacia de los tratamientos antitumorales existentes.

Para llevar a cabo la endocitosis del receptor, se requiere de la formación de un complejo compuesto por endofilinas, la ligasa de ubiquitina Cbl y la proteína adaptadora CIN85. Las proteinas Cbl conforman una familia ligasas de ubiquitina que unen a receptores de membrana entre los cuales se encuentra el EGFR, y está compuesta en mamíferos por c-Cbl o Cbl propiamente dicho, Cbl-b y Cbl-3. Por otro lado, las moléculas adaptadoras son polipéptidos generalmente no catalíticos que contienen uno o más dominios capaces de unir a otros ligandos tanto proteicos como no proteicos. Estas moléculas controlan de manera selectiva el ensamblaje espacial y temporal de complejos multiproteicos que transmiten señales intracelulares involucradas en la regulación del crecimiento, la diferenciación, la

migración y la supervivencia celulares. CIN85, junto con su homólogo CMS, pertenece a una familia de moléculas adaptadoras que contienen tres dominios SH3, una región rica en prolinas y un dominio *coiled-coil*.

La proteína adaptadora CIN85 interacciona con Cbl mediante la unión de dominios SH3 (Src homology 3) a la región carboxi-terminal de Cbl. Todos los dominios SH3 comparten un plegamiento conservado, en una conformación de sandwich o barril-β constituido por dos láminas-β antiparalelas. Dentro de esta estructura se encuentran dos lazos o *loops* variables, conocidos por lazos *RT* y *n-Src*. Una de las caras de este sandwich contiene una superficie de unión al ligando relativamente plana y de carácter hidrofóbico, que consta de 3 bolsillos poco profundos definidos por residuos aromáticos conservados. Los dominios SH3 reconocen generalmente ligandos ricos en prolinas, de secuencia consenso PxxP. Los grupos xP del motivo ocupan dos de los bolsillos hidrofóbicos formados por residuos conservados en la mayoría de los SH3. Existe un tercer bolsillo de unión el cual puede hospedar a una cadena lateral cargada positivamente, generalmente arginina o lisina, que flanquea el motivo de unión.

El ligando adopta una conformación helicoidal levógira extendida denominada como hélice poliprolina-2 (o PPII). El reconocimiento de esta estructura se consigue por la inserción de los extremos de la hélice PPII en pares de surcos complementarios en la superficie del SH3. Estos surcos se definen por una serie de residuos aromáticos conservados dispuestos de manera casi paralela.

Dependiendo de su orientación relativa respecto al complejo que formen con el dominio SH3, los ligandos de poliprolinas han sido clasificados como clase I o clase II. Los ligandos de clase I se unen con su extremo amino terminal al denominado lazo RT del dominio SH3, mientras que los ligandos de clase II se unen con su extremo carboxilo terminal al lazo RT. Los péptidos que se unen en una orientación de tipo I se ajustan al consenso +x*φ*Px*φ*P (clase I) (donde por lo general + es un residuo básico y *φ* es un residuo hidrofóbico), mientras que los péptidos que se caracterizan por *φ*P*x**φ*Px+ unen en la orientación opuesta. Sin embargo, la gran cantidad de dominios SH3 identificados hasta la fecha, ha permitido el amplio estudio de su diversidad estructural y funcional, seguramente por lo que se ha podido identificar varios motivos de unión que no corresponden al consenso.

Se ha determinado en Cbl una región mínima de unión a CIN85 ausente en Cbl-3. La unión de CIN85 a Cbl/Cbl-b parece estar mediada de forma independiente a motivos PxxP, siendo en este caso el motivo de reconocimiento un nuevo motivo poliprolina-arginina (PxxxPR), ya que las mutaciones de todas las secuencias PxxP presentes en la región C-terminal del dominio mínimo de unión no impedían la unión a CIN85.

Con la resolución de la estructura del primer dominio SH3 de CIN85 (CIN85-SH3A) en complejo con un péptido derivado de la secuencia natural de Cbl-b que incluye el motivo atípico de unión PxxxPR, se ha puesto de manifiesto un novedoso tipo de reconocimiento para esta clase de dominios. El complejo muestra una estructura heterotrimérica compuesta de dos dominios SH3 unidos simultáneamente al mismo péptido en orientaciones opuestas. El estudio estructural de dicho complejo ternario permitió identificar residuos fundamentales para tanto la

formación del heterotrímero como para la endocitosis y degradación del receptor EGFR. Este tipo de reconocimiento supone un nivel inesperado de oligomerización necesario para la regulación negativa del receptor.

Además, se ha podido resolver la estructura del SH3 homólogo en el adaptador de la misma familia, CMS, en complejo con el mismo péptido derivado de Cbl-b. Esto ha permitido revelar que la formación de un heterotrímero, similar en líneas generales, supone un reconocimiento específico del motivo de unión diferente al que ocurre en CIN85. El análisis detallado de las interacciones esenciales para la formación de dichos complejos triméricos, muestra diferentes residuos fundamentales en Cbl-b si la unión es a CIN85 o a CMS.

Estos hallazgos han permitido predecir el modo de unión a otros ligandos naturales de CIN85. La resolución del complejo del mismo dominio SH3 unido a un péptido derivado de CD2, confirmó aquellos residuos necesarios para la heterotrimerización, los cuales habían sido identificados previamente en el complejo con Cbl-b.

Asímismo, en el presente estudio se pretende descifrar las características esenciales que modulan este tipo de reconocimiento mediante un estudio comparativo con estructuras similares resueltas posteriormente.

INTRODUCTION

“The real purpose of introductions, of course, is to cite your own work, the work of your advisor, the work of your spouse, the work of a friend from college, or even the work of someone you've never met, as long as your name happens to be on the paper.”

E. R. Schulman

II. Introduction

Cells, as the smallest metabolically functional unit of life, need to perform essential functions for survival and development. "*Cells are required to stick precisely to the point. Any ambiguity, any tendency to wander from the matter at hand, will introduce grave hazards for the cells, and even more for the host in which they live*" (Lewis Thomas) (Thomas 1974).

Therefore, a rigorous control of cell morphology, growth and homeostasis, is vital for a correct function and development of tissues in multicellular organisms, and this depends on a very stringent regulation of responses to extracellular signals that modulate the level of cell activation. As a result of the binding of extracellular factors or ligands to their specific receptors, the formation of a broad repertoire of highly dynamic protein complexes takes place. The signal is thus amplified and translated into changes in the cellular proteome and genome (Crosetto, Tikkanen et al. 2005).

Most of the plasma membrane receptors elicit *positive signals*, and variations in the strength and duration of such signals convey different instructions to the cell. To ensure that proper signal thresholds are reached and maintained for the appropriate length of time, cells have developed specific *negative signaling* mechanisms. A delicate balance between positive and negative signals needs to be established and its deregulation is often implicated in the development of human diseases. An excess of positive signals that lead to excessive cell stimulation is found in cancers and autoimmune diseases, whereas dominance of negative signals is associated with various immune deficiencies (Dikic and Giordano 2003). Consequently, regulation of signal activation and inactivation plays a critical role in cell survival and development.

In order to attenuate the signal transmission, a number of highly complex chains of events need to occur in the cell. Negative signaling can take place transiently by regulating the strength and duration of the signal through multiple mechanisms such as steric hindrance, effector compartmentalization, posttranslational modifications (PTMs) or inhibition of catalytic activity (Dikic and Giordano 2003). However, when the signal needs to be promptly blocked, the signal inhibition must be irreversible.

Nature has consequently evolved a clever mechanism that ensures a rapid signal termination of transmembrane-receptor-triggered signal transduction. Membrane proteins cannot be redistributed in the cytosol, so that their internalization and posterior degradation are necessary for the regulation of their quantity present in the cellular membrane. Thus, to achieve prolonged cell desensitization to external stimuli, such as growth factors, endocytosis of cellular receptors takes place, and internalized receptors are then recycled or degraded.

1. Epidermal Growth Factor Receptor (EGFR) negative regulation

The Epidermal Growth Factor Receptor (EGFR) belongs to the family of Receptor Tyrosine Kinases (RTKs) and, together with its natural ligand (the Epidermal Growth Factor, EGF) has been identified as a key component of a wide variety of important biological processes, such as cell proliferation, differentiation, survival, migration, adhesion or apoptosis (Aaronson 1991; Schlessinger 2002; Dikic 2003).

As mentioned above, excessive cell stimulation by a prevalence of positive signals leads to cancer and autoimmune diseases. Constitutive activation of Receptor Tyrosine Kinases (RTKs) can take place by several mechanisms, among which are gene amplification, overexpression, chromosomal translocation or mutations (Finkel 2003; Normanno, De Luca et al. 2006). Therefore, it is not surprising that overexpression of EGFR is frequently found in colon, breast, bladder, cervical, ovarian, kidney, pancreatic and non-small cell lung cancer and it occurs with very high incidence in squamous cell carcinomas of the head and neck (Salomon, Brandt et al. 1995; Blume-Jensen and Hunter 2001; Schlessinger 2002; Finkel 2003; Normanno, De Luca et al. 2006). Mutations that activate the kinase activity of EGFR in the absence of ligand binding are as well often the main cause of tumorigenic changes in EGFR activity (Finkel 2003). The best characterized EGFR mutation is EFGRvIII, which includes deletions of exons 2 to 7 with consequent loss of residues 6 to 276. Such mutation bears a constitutively active receptor that is not regulated by endocitosis and is highly oncogenic (Nishikawa, Ji et al. 1994).

The establishment in 1984 of the relation between human cancer and the viral oncogene v-Erb-B, a constitutively active truncated mutant of Erb2 (an EGFR paralog), shed light to the knowledge that regulation of growth factor receptor activation is crucial to avoid aberrant signaling critically involved in human neoplasias (Downward, Yarden et al. 1984). Therefore, development of tyrosine kinase inhibitors directed against EGFR have been widely sought and used for treatment of certain cancers as advanced non-small-cell lung cancer (NSCLC) and other aerodigestive carcinomas (ADCs), the leading cause of cancer related death worldwide (Giaccone and Rodriguez 2005; Karamouzis, Grandis et al. 2007).

Down-regulation of EGFR, is hence a critical step in modulating receptor activity. It has been shown that interruption of receptor endocytosis alters duration and specificity of events that take place in this process and it has been proposed that breakdowns in this process might be as well of oncogenic character (Peschard and Park 2003; Crosetto, Tikkanen et al. 2005). Therefore, the study of the nature of events that ensure a proper modulation of duration, specificity and strength of receptor signal transduction stands out as a good target for cancer treatment.

This receptor down-regulation is mediated by endocytosis, receptor ubiquitination, and subsequent lysosomal degradation (Waterman and Yarden 2001). Key to this endocytosis is the formation of large protein complexes (Dikic and Giordano 2003) capable of removing activated receptors from the plasma membrane via clathrin-dependent or -independent pathways (Vieira, Lamaze et al. 1996; Soubeyran, Kowanetz et al. 2002; Bonifacino and Traub 2003; Szymkiewicz,

Shupliakov et al. 2004; Le Roy and Wrana 2005; Sigismund, Woelk et al. 2005). In this process, c-Cbl and Cbl-b proteins play a crucial role by acting as ubiquitin ligases as well as endocytic adaptor molecules (Sawasdikosol, Pratt et al. 2000).

Adaptor molecules are polypeptides with one or more domains able to bind other proteic or non-proteic ligands and, in general (although not the case of Cbl proteins), without enzymatic activity. They have the capability to connect proteins to other proteins, to the plasma membrane or to intracellular organelles. This is how large signaling multimeric complexes are formed, localizing signaling molecules to specific subcellular localizations, and hence contributing to the specificity and efficiency of cellular responses in a spatiotemporal manner (Pawson 2002; Csiszar 2006).

The same adaptors can act at different levels in different signalling cascades. This requires lower numbers of molecules, saving the metabolic cost that their delivery would suppose to the cell. More importantly, the *multi-tasking* capacity of adaptors allows dynamic signaling networks and integration of independent signaling events to happen through a limited number of unique proteins.

The negative regulation of the LET-23 receptor by SLI-1 (EGFR and Cbl orthologs in *C. elegans*) was the first example described of the regulation of RTKs ((Ryan, Davies et al. 2006) from (Langdon 1995)). Upon cell stimulation, Cbl proteins get translocated to activated RTKs and act as multifunctional adaptor proteins. As adaptors, Cbl proteins are capable of recruiting additional endocytic regulatory proteins such as CIN85(SH3KBP1)–endophilin complexes, which facilitate receptor endocytosis. Once the receptor is stimulated upon ligand binding receptor homodimerization and heterodimerization with receptors from the same family are triggered. The receptor gets phosphorylated and Cbl can then bind to specific phosphotyrosine motifs through its SH2 domain. Subsequently, Cbl multiple monoubiquitination of RTKs takes place and sorts the receptor-containing vesicle to a degradation route that leads to termination of the RTK signal (Dikic 2002; Haglund, Di Fiore et al. 2003; Haglund, Sigismund et al. 2003; Mosesson, Shtiegman et al. 2003).

Cbl proteins are polypeptides from a highly conserved family that control multiple essential cellular processes such as cell proliferation, differentiation and cell morphology (Dikic, Szymkiewicz et al. 2003; Schmidt and Dikic 2005). The overall domain structure of these proteins consists of an N-terminal Tyrosin Kinase Binding (TKB) domain, a Ring Finger motif, a Proline Rich region and a C-terminal UBiquitin Associated (UBA) domain that overlaps with a Leucine Zipper (LZ) motif (see Figure II.1). There are three described mammalian isoforms: Cbl-c (or Cbl-3), c-Cbl and Cbl-b. The first one is the shortest, lacking the C-terminal UBA/LZ domain and the CIN8/CMS binding region, and it is mainly expressed in epithelial cells. The other two are ubiquitously expressed, and hold a high degree of identity (47.4% overall sequence identity), however, c-Cbl contains some phosphotyrosine and phosphoserine regulatory sites that are absent in Cbl-b (Schmidt and Dikic 2005).

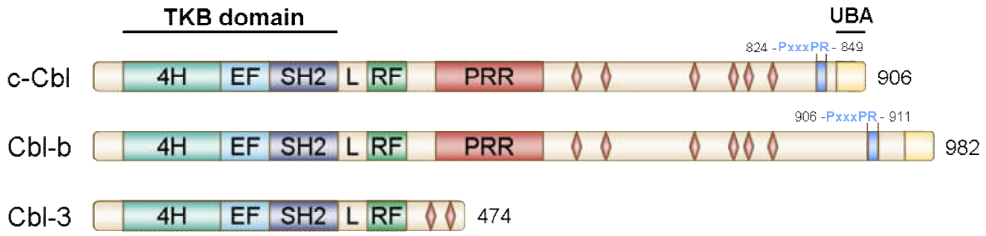


Figure II.1. Schematic representation of the domain structure of human Cbl proteins. All Cbl proteins share a high level of sequence conservation between their Tyrosine Kinase Binding (TKB) domain (which consists of a four-helix bundle (4H), a calcium-binding EF hand and a variant SH2 domain), their linker (L) region and RING finger (RF) domains. c-Cbl and Cbl-b have extensive Proline-Rich Regions (PRR) in their middle section that mediate interactions with numerous SH3-domain-containing proteins. In the C-terminus of c-Cbl and Cbl-b there is a PxxxPR motif that binds the SH3 domains from the CIN85/CMS family of adaptor proteins (residues 824-849 in c-Cbl and 906-911 in Cbl-b). Both c-Cbl and Cbl-b ligases have a UBA domain at the C-terminus that is lacking in the short Cbl-3. Brown rhombuses represent phosphorylation sites. Modified from (Thien and Langdon 2001).

1.1. The family of adaptor proteins CIN85/CMS

The recruitment of the adaptor molecule CIN85 by Cbl is crucial for the initiation of the early steps of the multistep process of EGFR downregulation (Dikic 2002; Petrelli, Gilestro et al. 2002; Soubeyran, Kowanetz et al. 2002). It was shown that inhibition of the Cbl-CIN85-endophilin interaction was sufficient to block EGFR endocytosis without affecting the ability of Cbl to ubiquitinate activated receptors (Soubeyran, Kowanetz et al. 2002).

A grown interest in the study of this protein has arisen with discoveries on the implication of CIN85 in the regulation of important processes such as cell apoptosis, cytoskeletal arquitecture or receptor down-regulation (Chen, Borinstein et al. 2000; Narita, Amano et al. 2001).

CIN85 (*Cbl-Interacting protein of 85 kDa*) was identified independently in different laboratories, and it is known with different names such as Ruk (*Regulator of ubiquitous kinase*) and SETA (*SH3 domain-containing gene Expressed in Tumorigenic Astrocytes*) in rat, and SH3KBP1 (*SH3-domain Kinase Binding Protein 1*) in mouse. Its homolog CMS (*Cas ligand with Multiple SH3 domains*) (Kirsch, Georgescu et al. 1999), also known as CD2AP (*CD2 Associated Protein*), shares an identical overall domain organization and high sequence identity with CIN85, and hence, it has been assumed to belong to the same family of adaptors and elicit similar biological functions.

Both CIN85 and CMS are ubiquitously expressed, although with slight, but important differences in their pattern of expression. Largest CIN85 expression levels were found in skeletal muscle, and lowest in lung and pancreas (Take, Watanabe et al. 2000). In contrast, CMS was observed to be expressed in high levels in the pancreas, placenta, colon, kidney and thymus, and low expression levels were detected in skeletal muscle, the aorta, bladder and uterus (Kirsch, Georgescu et al. 1999). This might suggest an overlapping role of these proteins with a specialized location for each of them.

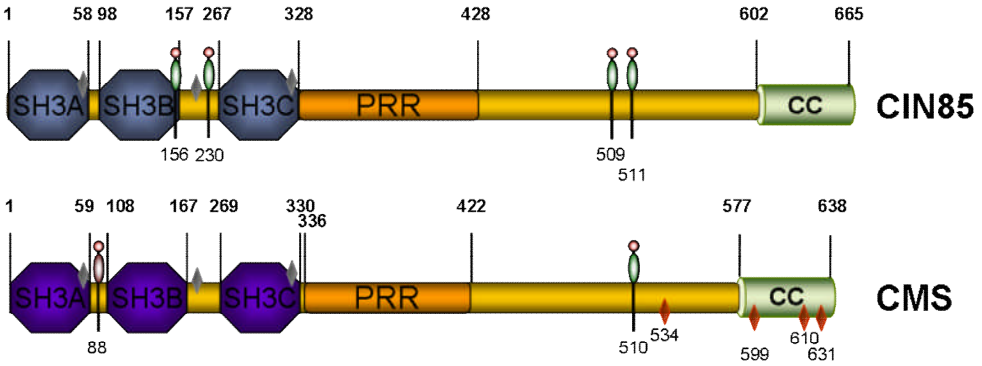


Figure II.2. Schematic representation of the domain structure of CIN85 and CMS. CIN85 and CMS contain three N-terminal SH3 domains, a centrally located proline-rich region (PRR) and a C-terminal coiled-coil (CC) domain. Boundaries of each domain are indicated with corresponding residue numbers. Four phophoserine sites are shown for CIN85 (Ser156 –putative-, Ser230, Ser509 and Ser511) and a phosphotyrosine (Tyr88) and a phosphoserine (Ser510) for CMS. FxDxF motifs and putative actin binding sites are indicated with grey and red rhombuses, respectively.

In the domain structure prediction of both adaptors there are three Src homology 3 (SH3) domains in the N-terminal end (SH3A, SH3B and SH3C), a central Proline-Rich Region (PRR), and a Coiled-Coil (CC) domain at the C-terminal end through which they are capable homo- and heterodimerizing and hence promote clustering (see Figure II.2). A region rich in serine and threonine residues, which could be subject to phosphorylation, lies between the second and third SH3 domains. There are also three FxDxF sequences in the amino-terminus that may serve as binding sites for the clathrin adaptor protein AP2 (Brett, Traub et al. 2002). The only apparent difference in CMS sequence, compared to CIN85, is the presence of four putative actin binding sites in its carboxyl-terminus that are similar to the LKKTET motifs found in a number of actin binding proteins (Kirsch, Georgescu et al. 1999).

The SH3 domains of these adaptors share higher similarity among themselves than to any other SH3 domains, suggesting that they may have overlapping specificities in binding. However, both CIN85 and CMS have been reported to display differential SH3 domain usage in interactions with the different binding partners which, suggested antagonistic activities for these adaptors (Tibaldi and Reinherz 2003).

Complex formation between CIN85 and Cbl occurs via the selective binding of SH3 domains of CIN85 to an atypical proline-arginine motif. This atypical motif is present in the C-terminal region of c-Cbl and Cbl-b, outside their Proline-Rich Region (PRR), and in other CIN85 effectors such as CD2 receptors (Dustin, Olszowy et al. 1998; Borinstein, Hyatt et al. 2000), B-cell linker protein BLNK (Watanabe, Take et al. 2000), apoptosis regulator AIP1 (Chen, Borinstein et al. 2000) and SETA binding protein SB1 (Borinstein, Hyatt et al. 2000; Take, Watanabe et al. 2000; Kowanetz, Szymkiewicz et al. 2003; Kurakin, Wu et al. 2003). Interestingly, it has been reported that deletion of the C-terminal end renders Cbl oncogenic (Lupher, Andoniou et al. 1998).

The closest homolog of CIN85-SH3A outside of the CIN85/CMS protein family is β-PIX, which has 52% sequence identity. Recently, β-PIX/Cool-1 has been described as another Cbl-interacting protein that modulates EGFR downregulation[18]. It was shown that activated Cdc42–PIX complexes promote sequestration of Cbl from the EGFR and consequently block EGFR endocytosis[19]. The crystal structure of β-PIX in complex with a Cbl-b derived peptide and binding experiments show that β-PIX binds to the same proline-arginine motif in Cbl that is recognized by CIN85 and hence may compete with CIN85 for binding to Cbl.

1.2. Molecular insights into the SH3 domain-mediated interactions

The human proteome is estimated to contain Src homology 3 (SH3) domains spread among over 600 proteins, according to the EBI InterPro database {http://www.ebi.ac.uk/interpro/IEntry?ac=IPR001452}, some of which bear more than one copy of SH3 domains (as seen for CIN85 and CMS). SH3 domains bind to proline-rich recognition sequences and are protein modules of general importance in many signalling processes of all organisms, suggesting that they regulate basic aspects of cellular organization.

Most SH3 domain-mediated physiological interactions are of moderate affinities. This facilitates the rapid response to activation or attenuation of cellular signals by the formation and dissociation of specific signaling complexes. Therefore, most protein-protein interactions occurring in cellular signal transduction are transient, making it a very dynamic process (Jia, Nie et al. 2005).

The first SH3 3D structure was reported in 1992 and since then, several hundred have been solved. They are generally composed of ~60-70 amino acids in a β-barrel structure made up of five antiparallel β-strands, (generally named β_a through β_e or β_1 through β_5; see Figure II.3). These β-strands form two β-sheets that pack against each other at approximately right angles. β-strand$_b$ has a torsion in the middle of about 90º known as a β-bulge, which allows it to participate in both β-sheets simultaneously. As a result, the first β-sheet is made up of β-strands *a*, *e* and the first half of *b*, and the second is composed of β-strands *c*, *d* and the second half of *b*. There are three variable loops: the RT-loop, which owes its name to its high content in arginine and threonine residues (located between β-strands *a* and *b*); the n-Src loop, named after the first protein in which the SH3 3D structure was described (between β-strands *b* and *c*); and the Distal loop, which is located in the opposite site of the ligand binding interface (between β-strands *c* and *d*). The variability in the length of the SH3 domains results generally from insertions or deletions in the n-Src and Distal loops. The three residues separating β-strands *d* and *e* are generally found in a 3_{10} helix conformation (Larson and Davidson 2000; Kishan, Newcomer et al. 2001).

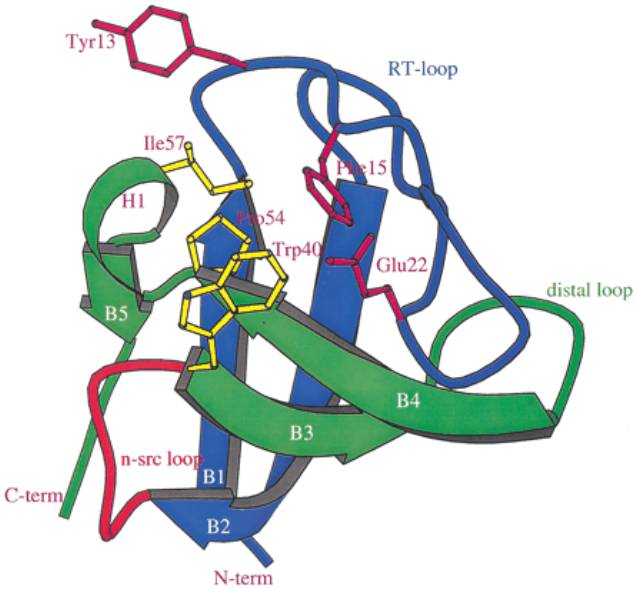


Figure II.3. "Schematic diagram of SH3 fold. The β-strands are labeled as B1 to B5 and are color coded (blue and green), indicating the strands separated because of swapping. The N terminus, C terminus, RT-loop, n-src loop, and distal loop are marked; the 3_{10}-helix is labeled as H1. The conserved residues responsible for polyproline binding are shown and labeled." Taken from (Kishan, Newcomer et al. 2001).

Larson and Davidson identified several positions in the amino acid sequences of 19 SH3 domains for which the structure was available at the time (2000) that were highly conserved and constituted the hydrophobic core that provides the protein domains with stability. Such hydrophobic core residues are predominantly inaccessible to the solvent, given that they are packed in the interior of the protein. These residues correspond in the first SH3 domain of CIN85 (SH3A) to positions 4 (usually Val), 6 (frequently Ala), 18 (Leu), 20 (generally Phe), 26 (Ile), 28 (typically Val), 37 (Trp), 39 (Gly), 48 (Phe) and 53 (Val).

This arrangement forms a relatively flat and hydrophobic surface, consisting of one acidic and two hydrophobic shallow pockets between the two β-sheets. This surface is involved in the recognition and binding of proline-rich motifs (Larson and Davidson 2000; Cesareni, Panni et al. 2002; Ferreon and Hilser 2004). The majority of SH3 domains bind to a proline-rich sequence with a φPxφP motif (more commonly denoted PxxP), where φ is usually a hydrophobic residue and x any residue (Mayer 2001). The ligand adopts an extended, left-handed helical conformation termed the polyproline type II (PPII) helix. Proline residues are in *trans* conformation, and the helical structure shows precisely three residues per turn (Adzhubei and Sternberg 1993) which makes this structure roughly triangular in cross-section, and the base of this triangle sits on the surface of the SH3 domain. The two hydrophobic ligand-binding pockets of the SH3 domain are occupied by two hydrophobic-proline (φP) dipeptides, whereas the acidic pocket typically interacts with a basic residue in the ligand distal to the φPxφP core (see Figure II.4). This basic residue binds the acidic pocket in the SH3 domain, which has been thought to provide the binding specificity and denoted as the "specificity pocket".

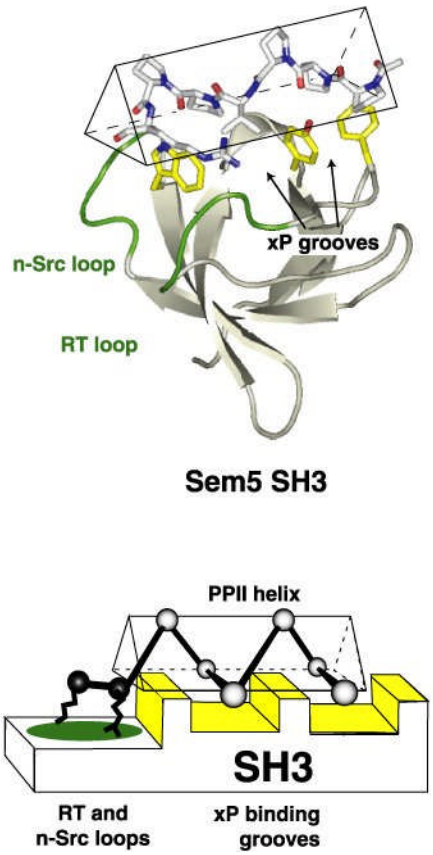


Figure II.4. Structure and binding mechanism of SH3 domains. The structure of the Sem5 SH3 domain in complex with a proline-rich ligand is shown. A cartoon of the proline-binding surface of these domains docked with a ligand, showing the general mechanism of recognition, is shown below. The core recognition surface has two xP binding grooves formed by aromatic amino acids, shown in yellow, and the adjacent, less conserved specificity pocket is designated in green." The polyproline peptide adopts a left-handed helical conformation of a PPII helix with a triangular cross-section (illustrated by the prism formed by connecting peptide residues parallel to the helix axis). Taken from (Zarrinpar, Bhattacharyya et al. 2003).

Proline-rich peptide ligands potentially bind a given SH3 domain in either one of two opposite directions governed by the location of such positively charged residue, usually Arg, which often precedes or follows the PxxP core element (Feng, Chen et al. 1994; Lim, Richards et al. 1994). According to this, the φPxφP (or PxxP) motif is usually further classified into +xφPxφP (class I) and φPxφPx+ (class II) (where + is generally a basic residue, usually arginine). Thereby, the φP dipeptides occupy different positions on the surface of the hydrophobic grooves, depending on the orientation of the ligand: this so that, relatively to the helix axis, class I binds conserved prolines in the motif in the same side as the basic residue, while class II does so in the opposite side (see Figure II.5).

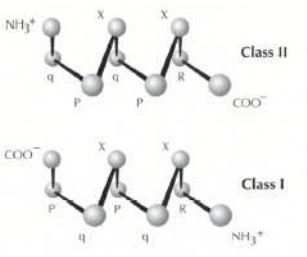


Figure II.5. "The structures of class I and class II peptide ligands are depicted schematically. P and R indicate conserved proline and arginine residues; q and X indicate variable amino acids", q being usually a hydrophobic residue. From (Mayer and Eck 1995).

Certain positions involved in peptide binding have been shown to be highly conserved among the SH3 domain sequences and to play an important role in binding preference. For instance, Musacchio and co-workers identified three amino acid substitutions in the mouse Abl SH3 domain that reversed binding specificity towards two proline-peptides (Musacchio, Saraste et al. 1994).

Structural analyses of different SH3-peptide complexes have helped identify certain positions in the SH3 sequences that show preference to selected amino acids. Almost totally conserved residues involved in peptide binding are amino acids corresponding to Pro49 and Trp36 in CIN85. Furthermore, extreme conservation observed at position of Tyr10 in CIN85 is explained by its location at the center of the peptide binding cleft. Residues corresponding in CIN85 to positions 8 (Phe or Tyr (Cesareni, Panni et al. 2002)), 35 (Gly), 51 (Asn) and 52 (Phe or Tyr (Cesareni, Panni et al. 2002)) were found to have low positional entropy values (Larson and Davidson 2000). Moreover, residues at the RT-loop, which, together with the n-Src loop, forms the acidic pocket, have been considered to have a role in the specificity of the binding, and positions 16 and 17 (CIN85 Asp16 and Glu17) are usually observed to be acidic residues (75% of the time according to Larson and Davidson's study) (Larson and Davidson 2000). These acidic residues typically hydrogen bond the arginine residue in the peptide motif, and hence explains the preference for acidic residues at this position (Musacchio, Saraste et al. 1994).

In spite of the fact that the shallow binding surface of an SH3 domain bears certain characteristics for the accommodation of the PPII helical structure, there are no intricate features that allow this binding surface to distinguish subtle differences between two proline-rich sequences. As a matter of fact, it has been shown that a given SH3 domain can interact with a few to several dozens of different peptide ligands (Li 2005). It is common in SH3-interacting proteins to contain multiple weak to moderate binding sites. Computational methods have indeed been developed to identify these putative canonical SH3 domain binding sequences within proline-rich regions of such proteins to narrow down the size of peptide libraries used for highthroughput analysis (as the recently published web server SH3-Hunter (Ferraro, Peluso et al. 2007)) The SH3 domain may dissociate quickly from one site and can subsequently be recaptured by a neighboring site in the same molecule. This way, the presence of multiple binding sites together with the capability of SH3 domains to recognize a collection of sequence motifs, effectively increases the local concentration of the SH3 domain and thereby promotes binding (Jia, Nie et al. 2005). Therefore, the incidence of various SH3 binding motifs in the same molecule may be a mechanism used by the cell to regulate SH3-mediated interactions (Li 2005).

Despite these overlapping recognition profiles, selectivity is essential in the SH3 domain-mediated interaction networks, given that the transduction of a defined signal from one protein to another requires a high level of integrity. Indeed, it was shown that substitution of the SH3 domain in Sho1 by other yeast SH3 domain avoided binding to its proline-rich target peptide in Pbs2 (Panni, Dente et al. 2002). Differences in binding motifs containing the sequence PxxP have been shown to offer different levels of specificity. In particular, the basic residue of the motif as well as additional interactions outside the motif, have played an important role at enhancing specificity in SH3 domain-mediated interactions (Zarrinpar, Park et al. 2003).

The evolution of a non-canonical recognition mechanism has offered an alternate solution to the need of the binding diversification of SH3 domains to achieve the required specificity. Several SH3 domains recognize non-PxxP motifs, as it has been shown for the SH3 domains of Eps8, which recognizes PxxDY (Mongiovi, Romano et al. 1999); Gads, which recognizes RxxK (Berry, Nash et al. 2002); or Fus1, which recognizes Arg-Ser-rich sequences (Tong, Drees et al. 2002).

“Although SH3 domains share ~30% amino acid similarity, and a common fold, the ability to predict the peptide recognition specificity of any given SH3 domain has been challenging” (Panni, Dente et al. 2002). Several studies have attempted to identify particular features in the SH3 sequences that would explain their specificity and “despite tremendous progress made in the past decade (…), this age old family of domains continues to generate surprises” (Jia, Nie et al. 2005).

1.3. Identification of a novel CIN85/CMS SH3 interacting motif in Cbl proteins

CIN85 interaction with Cbl proteins was intriguing in that its SH3 domains can only recognize a novel proline-arginine motif (PxxxPR) present in the C-terminal region of Cbl and Cbl-b (20, 21). In contrast to other SH3 domains, that are capable of recognizing atypical binding motifs in addition to the canonical PxxP, CIN85 SH3 failed to bind any polyproline motif from the PRR region in Cbl proteins.

Mutagenic experiments revealed the importance to the proline and arginine residues flanking the PxxxPR motif. Such motif indicated no correspondence to the previously described atypical SH3 binding sequences. Most remarkably, CIN85 homolog CMS was reported to preferentially recognized this atypical polyproline motif as well, while other SH3 containing proteins, such as Grb2, Src, Crk, Abl, CAP, or ArgBP2 did not (Kowanetz, Szymkiewicz et al. 2003).

Additional cellular CIN85 and CMS interacting proteins, such as CD2 receptor (CD2), Pak, Disabled 1 and 2, SETA-binding protein 1 (SB-1), SLP-65/BLNK, or Alg2-interacting protein (AIP1), and ASAP, contain the consensus sequence PxxxPR (8, 20). However, although binding to either of these two adaptors takes place at the same site, each interaction seems to elicit divergent biological consequences (Tibaldi and Reinherz 2003).

Identification of the consensus binding motif for the CIN85/CMS SH3 domains and structural determination of their interactions would provide a molecular explanation for the mechanism by which CIN85 and CMS act as inducible scaffolding proteins.

OBJECTIVES

III.Objectives

The appropriate balance of positive and negative signals is essential for the cell to bring about the correct responses. The negative regulation of the epidermal growth factor receptor (EGFR) signaling has been shown to be particularly important.

In the past seven years, since its identification, numerous discoveries have demonstrated the importance of CIN85 in processes that are crucial for the cell survival and regulation. One of its main functions relays on its participation in the EGFR endocytosis. The recruitment of CIN85 by Cbl seems to be necessary for the initiation of the early steps of the EGFR downregulation.

The study of the interaction between CIN85 and the ubiquitin ligases c-Cbl and Cbl-b raises a great interest not only in its global context but in the detail of the minimum domains that participate in such interactions. It has been demonstrated that these interactions are governed by the binding of CIN85 SH3 domains to an atypical PxxP-independent SH3 binding motif. This may represent a novel recognition mode that might explain the biological implications in the CIN85/CMS signaling cascades in particular and in the SH3 repertoire of interactions in general.

Establishing the principles that govern ligand recognition by the SH3 domains is critical to the understanding of their diverse functions in the cell (Li 2005). The broad knowledge on SH3 domain architecture and its interactions would allow us to perform a detailed analysis of the interaction of the SH3 domains of the CIN85/CMS family of adaptors with the c-Cbl/Cbl-b ubiquitin ligases, as well as a meticulous comparison with proteins of similar structure and/or function.

With this premises, the following objectives were proposed for the present study:

- To study, by means of structural analysis, the interaction between cin85 and the cbl ubiquitin ligases, in order to contribute to the understanding of the mechanism that controls the endocytic and degradation machinery of the tyrosine kinase receptors.

- To describe at the structural level the uncharacterized interface involving SH3 domains from CIN85 and the atypical PxxxPR motif in Cbl proteins.

- To characterize the structural differences of CIN85 and CMS SH3:Cbl complexes that could shed light to their functional divergence.

- To analyze the recognition of the PxxxPR motif in other CIN85/CMS natural partners and compare it with the recognition of Cbl proteins.

MATERIALS AND METHODS

“Be aware that negative results can be just as important as positive results”

E. R. Schulman

“Un resultado negativo es un resultado”

J. Bravo

IV. Materials and Methods

Several authors “have identified protein solubility (Yee, Chang et al. 2002) and crystallization as the two major bottlenecks in the process towards the determination of protein structures by X-ray diffraction” (Goulding and Perry 2003), given that first, one third to one half of prokaryote, and probably more of eukaryote proteins, cannot be expressed in soluble form in bacteria (Edwards, Arrowsmith et al. 2000), and second, “there is no apparent relationship between a protein’s properties and the conditions under which it will crystallize”, in which case, “one must empirically sample the limitless number of crystallization conditions” (Goulding and Perry 2003).

Only “~15–20% of small (<50 kDa) non-membrane proteins will be suitable immediately for structural biology” (Edwards, Arrowsmith et al. 2000), and it is known that even subtle changes in amino acid sequence can dramatically affect protein solubility (Edwards, Arrowsmith et al. 2000); thus, a common strategy is to clone and express many different constructs of different lengths or orthologs of the same gene and select the clone with the best solubility properties. Therefore, in order to guarantee a certain degree of success, an important portion of the time needs to be spent in the design of such protein constructs and subsequently in the acquisition of highly pure and soluble protein.

In order to do so, the protein sequence is inspected and a series of protein constructs is designed. The DNA is cloned in different vectors for the expression system of choice and effort is put on attempts for expression and solubilization of the protein of interest for the crystallization trials.

1. Cloning

All the CIN85 DNA templates were obtained from DNA clones provided by Ivan Dikic.

Enzymes, buffers and reagents used for DNA synthesis by PCR, DNA restriction, DNA dephosphorylation, DNA ligation and DNA recombination are described in the following table (Table IV.1).

Table IV.1. Enzymes, buffers and reagents used for DNA cloning.

Product	Supplier	Product code
PCR		
Pfu Polymerase	Biotools	10.512
Pfu Polymerase Buffer	Biotools	10.502
Taq Polymerase	CNIO	N/A (1 u/ml Taq Polymerase)
Taq Polymerase Buffer	CNIO	N/A (10x concentrated)
dNTP Mix	Fermentas	R0192
DNA restriction		
Nde I	Fermentas	ER0582
Xho I	Fermentas	ER0691
Buffer R (Red)	Fermentas	BR5
Nde I	Takara	1161B
Xho I	Takara	1094S
Buffer K	Takara	Supplied with Enzymes
DNA restriction		
Nde I	NEB	R0111S
XhoI	NEB	R0146S
Buffer 4	NEB	B7004S
BSA	NEB	B9001S
Nhe I	Fermentas	ER0971
Buffer Tango (Yellow)	Fermentas	BY5
Nhe I	Takara	1162A
Buffer M	Takara	Supplied with Enzymes
DNA dephosphorylation		
SAP	Roche	1758250
Buffer	Roche	Supplied with Enzyme
DNA ligation		
T4 DNA ligase	Fermentas	EL0015
T4 DNA ligase Buffer	Fermentas	Supplied with Enzyme

1.1. DNA amplicon generation

The acquisition of the CIN85 SH3 expression clones was accomplished via the Polymerase Chain Reaction (PCR) using specific primers designed individually for each open reading frame. Reactions were performed in volumes of 50µl, with 10-50ng of DNA template, 0.5µM of each oligonucleotide primer (see Table IV.2), 0.2mM dNTP mix, 1u high fidelity Pfu polymerase, and 1x Pfu polymerase buffer (see Table IV.1).

The oligonucleotides used for the generation of the different gene constructs either by PCR synthesis or by oligonucleotide hybridization were synthesized by Sigma-Proligo. The characteristics of the oligonuclotides utilized for the cloning of the gene constructs described here are summarized below (Table IV.2) (where the

restriction enzyme (RE) site sequences are shown in purple, as well as hanging bases for RE clamping and stop codons are shown in green lower case).

The melting temperature (T_m) shown in Table IV.2 was calculated for the hybridizing region (HR), following Wallace rule for short oligonucleotides (Wallace, Shaffer et al. 1979):

$$T_m = 4°C \cdot (\text{number of G's and C's}) + 2°C \cdot (\text{number of A's and T's})$$

For the full sequence (FS) the web-based Sigma-Proligo oligo calculation tool was used, for which calculation was based on the *Nearest-Neighbor* formula with assumed salt concentration of 50mM and single strand concentration of 250pM (Breslauer, Frank et al. 1986):

$$Tm = [\Delta H/((A + \Delta S) + R\ln (Ct/4))] + 273.15 + 16.6 \log [salt]$$

where **ΔH** (cal $mole^{-1}$) is the sum of the nearest-neighbor enthalpy changes for hybrid formation (<0); **A** (cal $K^{-1}mole^{-1}$) is a constant for helix initiation which is equal to -10.8 cal $K^{-1}mole^{-1}$ for nonself-complementary sequences and = -12.4 for self-complementary sequences); **ΔS** (cal $K^{-1}mole^{-1}$) is the sum of the nearest-neighbor entropy changes for hybrid formation (<0), **R** is the molar gas constant (1.987 cal $K^{-1}mole^{-1}$) and **Ct** is the total molar concentration of strands when oligonucleotides are not self complementary or it is equal to 4 times this concentration in the case of self-complementary sequences.

Table IV.2. Oligonucleotides used for the amplicon generation of the CIN85 SH3 constructs

Name	Protein	RE site	Tm (°C)		Sequence 5'-3'
			FS	HR	
FwSH3A	$CIN85_{2-6}$(Fw)	NdeI	62	52	cgaCATATGGTGGAGGCCATAGTGG
RvSH3A	$CIN85_{53-58}$(Rv)	XhoI	61	44	gggccCTCGAGTTTCTTTATTTCTCTTAC
RvStopSH3A	$CIN85_{53-58}$(Rv)	XhoI	57	44	gggccCTCGAGtcaTTTCTTTATTTCTCTTAC
FwSH3B	$CIN85_{101-105}$(Fw)	NdeI	77	62	gcgaCATATGCGCCGGTGCCAGGTGGC
RvSH3B	$CIN85_{153-157}$(Rv)	XhoI	72	60	ccgCTCGAGCCCTGACAGCTCCTTGATG
RvStopSH3B	$CIN85_{153-157}$(Rv)	XhoI	68	60	ccgCTCGAGtcaCCCTGACAGCTCCTTGATG
FwSH3C	$CIN85_{270-276}$(Fw)	NheI	62	37	gcgGCTAGCGATTACTGCAAAGTAATATTT
RvSH3C007New	$CIN85_{324-328}$(Rv)	XhoI	65	48	cggCTCGAGCGGTGGAAGTAACTTC
RvSH3C DoubleStop	$CIN85_{324-328}$(Rv)	XhoI	61	48	gtcCTCGAGctattaCGGTGGAAGTAACTTC
Fwc-CblPeptide	$cCbl_{825-835}$	NdeI/ XhoI	91	91	gcgCATATG CCCGAGAGGCCTCCAAAACCATTCCCGCGGAGA tgaCTCGAGgcg
Rvc-CblPeptide	$cCbl_{825-835}$	NdeI/ XhoI	91	91	cgcCTCGAGtca TCTCCGCGGGAATGGTTTTGGAGGCCTCTCGGG CATATGcgc

Table IV.2. RE: Restriction Enzyme;**Tm**: Melting temperature; **FS**: Full Sequence; **HR**: Hybridizing Region.

Unless otherwise specified, the reactions were performed in a MJ Research MiniCycler PTC150 thermal cycler, with an initial three minute denaturing step at 95ºC, followed by 30 cycles of three steps (denaturing step: 30'' at 95ºC; annealing step: 45'' at a temperature 5ºC lower than the lowest primer melting temperature (not lower than 45ºC); extension step: 2' per kbp of DNA to be amplified (1' minimun) at 72ºC) and a final extension step of 10' at 72ºC.

In the case of the pGATEV/c-Cbl$_{820-830}$ clone, the dsDNA corresponding to the c-Cbl CIN85 binding region was obtained by denaturation at 95ºC and posterior annealing through slow cooling at room temperature (RT) of an equimolar mix of the long complementary oligonucleotides Fwc-CblPeptide and Rvc-CblPeptide (see Table IV.2).

1.2. DNA digestion and dephosphorylation

Amplified DNA and appropriate vector were digested with Nde I (Nhe I in the case of CIN85-SH3C) and Xho I DNA Restriction Enzymes following the manufacturer's recommendations. When possible, simultaneous double digestions were performed.

After digestion, in order to avoid re-circularization, the vector was dephosphorylated with 1u of Shrimp Alkaline Phosphatase (SAP) in each of two consecutive incubations of 30' at 37ºC, as specified by the manufacturer. DNA was subsequently purified as described in section IV.1.7.

1.3. DNA ligation

PCR and vector DNA concentrations were estimated by evaluation of fluorescence intensity under UV light in samples separated by agarose gel electrophoresis. Concentration ratios were calculated taking into account intensity and size of the DNA fragments (see Figure IV.1).

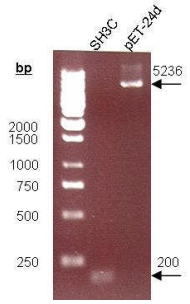


Figure IV.1. Example of intensity calculation for a vector and an insert used for DNA ligation. Same volume of Nhe I/Xho I cut pET-24d and CIN85-SH3C$_{270-328\ TAA\ TAG}$ were loaded in a 0.8% agarose-TBE gel. The intensity of the pET-24d vector is calculated to be 10 times larger than that of the CIN85-SH3C$_{270-328\ TAA\ TAG}$ insert. Given that the vector is 30 times longer than the instert, we calculate a 1:3 insert:vector concentration ratio.

Approximately 100-200ng of vector were used for each DNA reaction and different ratios of vector:insert DNA (1:5, 1:10, 1:15) were calculated for the 20μl reactions (when DNA insert was very diluted, 50μl reactions were prepared). 1u of T4 DNA ligase was used per reaction in 1x T4 ligase buffer at 22ºC. Ligation times

varied from 2 hours to o/n. A ligation reaction without DNA insert was also set as a negative control for posterior evaluation of vector re-circularization.

The characteristics of the plasmids used for the construction of the different gene constructs are described in the following table:

Table IV.3. *E.* coli expression vectors.

Plasmid	Antibiotic Resistance	Tag	Cleavage site	Size	Supplier (Cat. #)
pET-21a	Amp	N-ter T7 C-ter His_6	No	5.4 kb	Merk[1] (69740-3)
pET-29b	Kan	N-ter S C-ter His_6	Thrombin	5.4 kb	Merk[1] (69872-3)
pET-24d	Kan	N-ter T7 C-ter His_6	No	5.3 kb	Merk[1] (69752-3)
pGATEV	Amp	N-ter His_6 N-ter GST	Tev	4.7 kb	Vector modified at the Max-Plank-Institute for Molecular Physiology, Germany; and the Institute of Protein Research, Russia (Kalinin, Thoma et al. 2001)

1.4. DNA transformation

The ligation reactions were transformed into DH5α bacterial cells (see Table IV.4) previously treated to make them electro or chemically competent for their transformation with the different DNAs by electroporation or heat shock transformation procedures (for transformation of ligation reactions the latter procedure was preferred).

1.4.1. Preparation of transformation competent *E. coli* cells

DH5α is an *E. coli* strain with a high transformation efficiency that is recombination-deficient and thus very useful for plasmid DNA amplification. In order to make them electrocompetent, a 1L culture is grown overnight and cells are harvested by centrifugation at 4ºC and 4000xg for 20 minutes. They are then resuspended in 1L ice-cold 1mM Hepes pH7. The cells are washed again with 0.5L of Hepes and resuspended in 20ml 1mM Hepes pH7.0, 10% Glycerol for a final 20 minute centrifugation. They are finally resuspended in 3ml 10% Glycerol, frozen in dry ice in 50μl aliquots and stored at -80ºC.

To make cells chemically competent, Current Protocols procedure was followed (Seidman 1997), except that final cell pellet was resuspended in 2.5 times

[1] Previously Novagen

less the recommended volume, and aliquots of 200µl were made and subsequently frozen in dry ice prior to storage at -80ºC.

Both procedures were carried out under sterile conditions, and contamination controls were performed.

1.4.2. Transformation of *E. coli* cells by electroporation or heat shock

Transformation of electrocompetent cells was carried out using 25µl of, already thawed under ice, electrocompetent cells and 0.2µl of pure plasmid DNA (or 50µl of cells and 1µl of insert-plasmid ligation reaction) in a prechilled 0.2µl gap Gene Pulser/MicroPulser Cuvette from BioRad (Cat. # 165-2086). The Gene Pulser II BioRad machine is set to give a pulse of 25µF, 2.5kV with the pulse controller at 200ohms for 4msecs. Cells are quickly diluted into 1ml glucose containing SOC medium (section IV.2) and incubated at 37ºC for a 30 to 45 minute recovery period before plating into an antibiotic containing LB-Agar plate.

The procedure for transformation of chemically competent cells with plasmid DNA was performed by means of heat shock reaction. Either 60µl of $CaCl_2$ treated cells and 10ng of plasmid DNA or 90µl of cells and the whole ligation reaction are mixed and incubated on ice for 45 minutes or more. A 1.5 minute 42ºC heat shock is applied to DH5α cells (45-50 seconds for expression strains) and bacterial cells are incubated on ice for five more minutes. Cells are diluted with 200µl of LB medium and incubated at 37ºC for 30 minutes to 1 hour for recovery before plating them in LB-Agar plates with appropriate antibiotic.

1.5. Confirmation of the DNA sequence

A number of colonies obtained by transformation and plating of vector:insert ligation reactions into DH5α *E. coli* cells were selected and subjected to either colony PCR analysis, restriction enzyme digestion of the extracted plasmid DNA, or both.

Colony PCR was performed following the procedure described in section IV.1.1, except that, instead of high fidelity polymerase, Taq polymerase was used with its buffer and 2.5mM $MgCl_2$, and the colony was used as DNA template with an initial denaturing step of 5’ instead of 3’.

Plasmid DNAs extracted from positive hits were sequenced in the CNIO facilities, in a multicapillary DNA analyzer ABI Prism 3700 (Applied Biosystems) to confirm the integrity of the sequences.

1.6. DNA detection

1.6.1. DNA electrophoresis

For DNA detection, 0.8% Agarose-TBE (89mM Tris, 89mM Boric Acid, 1.9mM EDTA) with 0.5mg/ml Ethidium Bromide gels were prepared by dilution of the agarose in TBE buffer by boiling at 750W in a microwave oven. The Ethidium Bromide is added when solution is warm, and solution polymerization takes place at RT.

Electrophoresis of the DNA samples was performed at 20-90mA in TBE buffer and DNA size was monitored by comparison to 1kb and 100bp DNA ladders from Promega and Biotools, respectively (catalog numbers G5711 and 31.006).

Samples were prepared with 6x TriTrack^{TM} Loading Dye Solution from Fermentas (Cat. #R1161) or 2x home made Loading buffer (0.04% Bromophenol Blue (Sigma B-0126), 0.04% Xylene Cyanol FF, 10% Glycerol) to a >1x final concentration.

1.6.2. DNA quantification

The absorbance of the extracted DNA was measured at 230, 260 and 280nm wavelengths so that the purity could be evaluated with the double ratio method (260/230nm[2] and 260/280nm[3] absorbance ratios) that give indication of the presence of organic compounds or protein respectively. The concentration of the DNA samples were calculated from the absorbance at 260nm assuming 1.0 OD_{260} per 50ng/μl dsDNA.

1.7. DNA purification

DNA purification was achieved using different purification kits: for purification of PCR products, RE, and ligation reactions, JetQuick PCR Purification Spin kit (Cat. # 410050) was used; for plasmid DNA extraction from DH5α *E. coli* cells, Promega Wizard Plus SV Minipreps kit (Cat. #A1460) was used for minipreps and Qiagen Plasmid Midi/Maxi kits were used for large culture volumes; and for gel DNA extraction both Qiagen Gel Extraction Kit (Cat. #28704) and JetQuick Gel Extraction Spin Kit (Cat. #420050) were used indistinctively. In all cases, the manufacturers' protocols were followed, except that for Qiagen Midipreps, DNA was more extensively washed after precipitation (an extra washing step was performed with 1ml 96% Ethanol followed by a second washing step with 1.5ml 70% Ethanol).

---

[2] Monitored to be near 2.0 OD for high purity DNA and low organic compounds content (Warburg, O. and W. Christian (1942). "Isolierung und Kristallisation des Gärungsferments Enolase (Isolation and crystallization of enolase)." Biochem. Z. **310**: 384-421.).

[3] Monitored to be beyond 1.8 OD for high purity DNA and low protein content.

In all cases, DNA was eluted with 65-70ºC hot DNAse free water and column membrane was let stand for at least one minute for proper hydration, in order to achieve optimal DNA elution.

In some instances, standard DNA salt precipitation protocols were followed to purify low concentrated DNA or small amounts of DNA from PCR reactions. Such method consisted on precipitation of DNA with either $^{25}/_{3}$ DNA volumes of 3M NaCl or $^{1}/_{10}$ DNA volumes of 3M Sodium Acetate, and 2-2.5 volumes of 96% Ethanol which was left overnight at -20ºC (or incubated at -80ºC for 1-2hr). Precipitated DNA was centrifuged 10-30 minutes at maximum speed at 4ºC and supernatant was carefully removed by pipeting. DNA pellet was finally washed with 70% Ethanol, vortexed and centrifuged 15 minutes at 4ºC at maximum speed. Ethanol traces were evaporated by a 5 minute incubation at 37ºC and DNA was resuspended in 10μl water. Hydration of DNA was achieved by o/n incubation at 4ºC. DNA was subsequently quantified and stored at -20ºC.

2. Heterologous protein over-expression and solubilization

Once the sequences were inspected and no mutation was detected, the plasmid DNAs were transformed in different *E. coli* expression strains by electroporation or heat shock as described in section IV.1.4.2.

The bacterial *E. coli* strains used in this study are described in the following table (for convenience, the strain DH5α, used for DNA amplification and not protein expression, has been added to the same table).

Table IV.4. *E. coli* strains

Strain	Genotype	Comments	Supplier
DH5α	F^- recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 AlacU169Φ80d lacZAM15	Used for DNA amplification Highly transformation efficient of unmethilated DNA and recombination-deficient	Invitrogen
BL21 (DE3)	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Used for protein expression T7 promoter driven expression is repressed until IPTG induction.	Invitrogen
BL21 (DE3) pLys S	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS(cm^R)	Used for protein expression Reduces and almost eliminates expression from transformed T7 promoter containing plasmids when not induced.	Invitrogen
Rosetta (DE3) pLys S	F^- ompT hsdSB ($r_B^- m_B^-$) gal dcm (DE3) pLysS/RARE (argU, argW, ilex, glyT, leuW, proL) (cm^R)	Used for protein expression Expresses rare tRNAs facilitating expression of genes that encode rare *E.coli* codons.	Novagen

Strain	Genotype	Comments	Supplier
TG1tr	D(lac-pro) glnV44(AS) thi hsdD5 F0a	Used for protein expression T phage resistant.	Modified from TG1 strain at the University of Cambridge (Wang, Lo et al. 2005)

{http://openwetware.org/wiki/E._coli_genotypes#BL21.28DE3.29}

F$^-$ = Does not carry the F plasmid, **recA1** = For reduced occurrence of unwanted recombination in cloned DNA; cells UV sensitive, deficient in DNA repair, **endA1** = For cleaner preparations of DNA and better results in downstream applications due to the elimination of non-specific digestion by Endonuclease I, **gyrA96** = mutation in DNA gyrase; conveys nalidixic acid resistance, **gyrA462** = mutation in DNA gyrase; conveys resistance to ccdB colicin gene product, **thi** = requires thiamine, **hsdR** = For efficient transformation of cloned unmethylated DNA from PCR amplifications, **supE** = **glnV**: suppression of amber (UAG) stop codons by insertion of glutamine; required for some phage growth, **relA** = relaxed phenotype; permits RNA synthesis in absence of protein synthesis, **(φ80)** = Cell carries the lambdoid prophage φ80. A defective version of this phage carrying lacZM15 deletion is present in some strains, **lacZΔM15** = partial deletion of the lacZ gene that allows α complementation of the β-galactosidase gene; required for blue/white selection on XGal plates. Deletes the amino portion of lacZ (aa 11-41), **DE3** = T7 polymerase promoter, **ompT** = mutation in outer membrane protein protease VII, reducing proteolysis of expressed proteins, **gal** = galactose loss of function, **dcm** = cytosine methylation at second C of CCWGG sites abolished, **lon** = lon protease loss of function, **hsdS** = Both restriction and methylation of certain sequences is deleted from the strain. If you transform DNA from such a strain into a wild type strain, it will be degraded, **r_B^- m_B^-** = The B defines the strain lineage. The - indicates the strain hasn't got the restriction system (r_B^-) nor the modification (methylation) system (m_B^-), **lacI** = overproduction of the lac repressor protein, **pLysS** = contains pLysS plasmid carrying chloramphenicol resistance and phage T7 lysozyme, effective at attenuating activity of T7 RNA polymerase, for better inhibition of expression under non-induced conditions, **cm^R** = resistance to Chloroamphenicol.

Colonies obtained from DNA transformed *E. coli* cells were selected on LB-agar with proper antibiotic (LB medium with 4% agar). Selected colonies were subsequently grown o/n at 37ºC in LB medium (see below). Short scale, 20ml, expression tests were performed by diluting the o/n cultures 20-50 times in different expression strains, at different temperatures (16, 20, 30 and 37ºC).

Different rich media supplemented with appropriate antibiotic (100μg/ml Amp, 50μg/ml Kan or 34μg/ml Chloroamphenicol) were tested for protein expression: *Lysogeny Broth* (more commonly known as *Luria Broth*) (LB) (10g/L Tryptone, 5g/L Yeast extract, 170mM NaCl, pH7.2); *Two times Tryptone-Yeast extract* (2xTY) (16g/L Tryptone, 10g/L Yeast extract, 85mM NaCl, pH7.2); *SOC* (20g/L Tryptone, 5g/L Yeast extract, 8.55mM NaCl, 0.25mM KCl, 2M $MgCl_2$, 1M Glucose, pH7); and *Terrific Broth* (TB) (12g/L Tryptone, 24g/L Yeast extract, 0.4% Glycerol, 0.17M KH_2PO_4, 0.72M K_2HPO_4) (all these media were prepared by the CNIO services).

When the optical density at 600nm (OD_{600}) of the cell culture reached 0.6-1.0 (corrected for turbulence by a factor of 2), expression was induced with different concentrations of Isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.1-1mM. Samples were collected every two hours and protein samples were prepared as described in section IV.4.1 for SDS electrophoresis analysis.

Cell cultures were pelleted 15' at 6000xg and resuspended in different lysis buffers (see Table IV.5), sonicated in a Branson 250 Sonifier at a 10% intensity for 2' and centrifuged 15'-30' at 16000xg. Amount of soluble fraction was evaluated by gel electrophoresis of the supernatant (s/n) and the pellet of the cell lysates.

Table IV.5. Solubilization buffers used for cell lysis.

Solubilization/Lysis Buffers		
Name	**Composition**	**Proteins**
High [NaCl]	50mM Tris pH 7.5 at 4ºC, 400mM NaCl, 15mM Imidazole	CIN85-SH3B$_{101-157}$His$_6$
Medium [Imidazole]	50mM Tris pH 8.0 at 4ºC, 200mM NaCl, 50mM Imidazole	CIN85-SH3C$_{270-328}$His$_6$
PBS (Phosphate Buffer Saline solution)	137mM NaCl, 10mM Na_2HPO_4-2$H_2$0, 2.7mM KCl, 2mM KH_2PO_4 at pH 7.4	CIN85-SH3C$_{270-328\ TAA\ TAG}$:His/GST/cCbl$_{820-830}$
Low [NaCl]	50mM Tris pH 8.0 at 4ºC, 100mM NaCl, 15mM Imidazole	CIN85-SH3A$_{2-58}$His$_6$
Tris	100mM Tris pH 7.5, 200mM NaCl, 1mM EDTA	CIN85-SH3A$_{2-58\ tga}$, CMS-SH3A$_{1-62\ tga}$

Once a suitable amount of protein expression and solubilization was achieved, the expression protocol was established for large scale 1L cultures. A glycerol stock was frozen in dry ice (1.5ml of o/n culture was centrifuged at 5000xg at 4ºC and resuspended in 1ml 15% Glycerol-LB) and stored at -80ºC for posterior use for initiation of o/n cultures.

Large scale cultures were grown in incubator shakers (New Brunswick scientific Innova 4430) at the indicated temperature (see Results section) with a shaking speed of 200rpm. After standard times of expression of 4-6hr (o/n for expression at 16 or 20ºC), cells were harvested by pelleting in 1L bottles using a Beckman Coulter Avanti J-20 XP centrifuge in a JLA-8.1000 rotor for 15min at 5000xg. Total cell pellets were washed with PBS buffer (Table IV.5), frozen under liquid nitrogen, and stored at -80ºC until used for protein purification.

3. Protein purification

Thorough purification is very important for protein crystallization, as contaminants can hamper crystallization. Thus, multiple step purification was undertaken for each protein.

For the purification of the already expressed protein, appropriate total cell pellets corresponding to 1-3L cultures (depending on the expression level achieved), were thawed in ice and resuspended in 5-10ml of the indicated solubilization/lysis buffer per gram of cells.

Cell lysis was achived by mechanical methods such as homogenization (with an Avestin Emulsiflex-C5) or sonication. Sonication was the lysis standard method of choice and was carried out on ice, in 10-15 mins. with short pulses of 0.5 sec. with resting periods of 2 seconds in between, at 40% intensity in a Bioblock Scientific Vibracell 75042, which was located inside a 4ºC room.

Membrane traces were discarded by high speed centrifugation at 16100xg in a Beckman ultracentrifuge with a Ti 45 rotor at 4ºC for 40 minutes. The cleared lysate was filtered through a 0.45μm membrane prior to loading into a previously equilibrated affinity column.

3.1. His-tagged and GST-tagged protein purification

The purification protocol followed for His-tagged proteins consisted on the loading of the 4ºC or ice cold filtered proteins in a HiTrap GE Healthcare[4] column at 5ml/min with a MiniPuls 3 peristaltic pump (Gilson) or with an AKTA Purifier (GE Healthcare). The columns were previously equilibrated with the lysis buffer (with the NTA resin previously charged with 100mM $NiSO_4$).

Once the protein was bound to the column, the sample was washed with specific washing buffers at 5ml/min in an AKTA Purifier (GE Healthcare) and eluted in a 5-10 column volume gradient to 100% elution buffer, which contained 500mM Imidazole or 100mM EDTA (10mM L-Glutathione reduced (GSH) in the GST co-purification of the CIN85-SH3C$_{270\text{-}328\ TAA\ TAG}$:His/GST/cCbl$_{820\text{-}830}$ complex). A final 5 column volume wash was carried out to guarantee complete protein elution.

If the purity of the protein sample was not satisfactory, it was subjected to a second purification step in an anion exchanger sepharose column (HiTrap™ Q HP or FF, GE Healthcare Cat. #17-1154-01 and 17-5156-01, respectively). Sample was diluted to 50-100mM NaCl and/or 20mM Imidazole concentration, and loaded in the column to be eluted in a NaCl gradient.

Elute fractions were analyzed by SDS-gel electrophoresis (section IV.4.1) and pure fractions were subsequently concentrated by ultracentrifugation with Amicon Ultra or Vivaspin filter devices to a final volume of 1-2ml. The concentrated fraction was then centrifuged at ≥14000xg for 10-15 minutes to bring down any particle or aggregate present in solution and loaded into a HiLoad Superdex 75 16/60, Superdex 75 26/60 or Superdex 30 16/60 gel filtration column for a size exclusion separation, using an AKTA Prime system at at 0.5ml/min for the 16/60 and 2.5ml/min for the 26/60.

Purification steps and buffers used for each protein are specified in the following table.

[4] Previously Amersham. Columns: HiTrap™ Chelating HP (Cat. #17-0409-03), HisTrap™ HP or FF (Cat. # 17-5248-02 or 17-5255-01), GSTrap™ FF (Cat. #7-5131-02).

Table IV.6. Purification schemes for CIN85-SH3 constructs and CMS-SH3A.

Protein	Column	Wash buffer	Elution buffer	Superdex Buffer
CIN85-SH3B$_{101-157}$His$_6$	HiTrap Chelating	Wash 1: 50mM Tris pH 7.5 at 4ºC 400mM NaCl 15mM Imidazole Wash 2: 50mM Tris pH 8 at 4ºC 400mM NaCl 0.1% Triton x-100 v/v 2mM Imidazole	50mM Tris pH 8.0 at 4ºC 1% Ethylene Glycol v/v 500mM Imidazole	20mM Hepes pH 8.0 at 4ºC, 100mM NaCl, 4% Glycerol
	HiTrap Q FF	Wash: 20mM Hepes pH 8.0 at 4ºC	20mM Hepes pH 8.0 at 4ºC 1M NaCl	
CIN85-SH3C$_{270-328}$His$_6$	HiTrap Chelating	Wash: 50mM Tris pH 8.0 at 4ºC 200mM NaCl 50mM Imidazole	50mM Tris pH 8.0 at 4ºC 1% Ethylene Glycol v/v 500mM Imidazole	20mM Hepes pH 8.0 at 4ºC 100mM NaCl 4% Glycerol
	Q FF	Wash: 20mM Hepes pH 8.6 at 4ºC	20mM Hepes pH 8.0 at 4ºC 2M NaCl	
CIN85-SH3C$_{270-328\ TAA\ TAG}$: His/GST/cCbl$_{820-830}$	GSTrap FF	Wash 1: 50mM Tris pH 8.0 at 4ºC 500mM NaCl Wash 2: 50mM Tris pH 8.0 at 4ºC 100mM NaCl	50mM Tris pH 8.0 at 4ºC 100mM NaCl 10mM GSH	25mM Tris pH 8.0 at 4ºC
CIN85-SH3A$_{2-58}$His$_6$	HiTrap Chelating	Wash 1: 50mM Tris pH 8.0 at 4ºC 400mM NaCl 15mM Imidazole Wash 2: 50mM Tris pH 8.0 at 4ºC 2mM Imidazole	50mM Tris pH 8.0 at 4ºC 1% Ethylene Glycol v/v 100mM EDTA	20mM Hepes pH 7.0 at 4ºC 100mM NaCl 4% Glycerol
	Q HP	Wash: 50mM Hepes pH 8.0 at 4ºC	50mM Hepes pH 8.0 at 4ºC 500mM NaCl	
CIN85-SH3A$_{2-58\ tga}$ **CMS-SH3A$_{2-62}$**	HiTrap Phenyl FF (high sub)	Wash: 100mM Tris pH 7.5 200mM NaCl 1mM EDTA 50% $(NH_4)_2SO_4$	100mM Tris pH 7.5 200mM NaCl 1mM EDTA	20mM Tris pH 7.5 at 4ºC 150mM NaCl

3.2. Untagged CIN85 and CMS SH3A purification

For the purification of CIN85-SH3A$_{2\text{-}58\ tga}$ and CMS-SH3A$_{2\text{-}62}$, the total cell pellets of 3L cultures were resuspended with 5ml of Tris buffer (see Table IV.5) per gram of cells, sonicated and ultracentrifuged as described previously. The clear lysate proteins were precipitated with Amonium Sulfate in two steps at 4ºC for 15 minutes and centrifuged at 13000xg for 15 minutes. The first precipitation step was of 50% w/v Amonium Sulfate, followed by a second step in which the salt concentration was increased to 70% w/v Amonium Sulfate.

The soluble fraction was then diluted to 50% ammonium sulfate Tris buffer to avoid any trace of precipitated protein, and filtrated through a 0.45μm membrane. The sample was then loaded into a hydrophobic interaction chromatography HiTrap column from GE Healthcare (HiTrap Phenyl (high sub) FF, Cat. # 17-5193-01) (previously equilibrated with 50% $(NH_4)_2SO_4$/Tris buffer) and the protein was eluted with a 50 to 0% $(NH_4)_2SO_4$ gradient. Protein obtained from this step was very pure, so that it was concentrated with an Amicon 5kD filter to a volume between 1.5 and 2ml and loaded into a Superdex 75 16/60 column, which was run with 20mM Tris pH 8.0, 200mM NaCl at 0.5ml/min. To decrease the concentration time, the sample was diluted to a calculated Amonium Sulfate concentration of 25-50% prior to protein concentration.

4. Protein detection

4.1. SDS-PAGE

The purity and relative molecular mass of protein samples as well as the progress of the purification procedure was followed by separation of proteins by Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE).

SDS-polyacrylamide gel electrophoresis (PAGE) was performed according to the Laemmli method (Laemmli 1970) using gels of uniform or gradient density in a MiniProtean 3 system from BioRad.

Uniform density polyacrylamide gels were prepared as described in the "Maniati's Laboratory Manual" (Sambrook and Russell 2001). Polyacrylamide gradient gels were prepared with a Model 485 Gradient Former from BioRad, following BioRad's protocol with the following modifications: 41.5ml of the 4% polyacrylamide Light Solution, and 55ml of a 23% polyacrylamide Heavy Solution.

Protein samples were diluted with with 5x Protein Loading Sample Buffer (10% SDS, 312mM Tris pH 6.8, 50% Glycerol, 0.05% Bromophenol Blue (Sigma B-0126), 712.5mM β-Mercaptoethanol (β-MeOH)) and boiled at 99ºC for 3-10 minutes. In the case of total cell pellet (TCP) protein samples, the OD_{600} of the cell cultures were measured, 0.5ml of cell culture was pelleted and resuspended in 2x Protein Loading Sample Buffer (4% SDS, 125mM Tris pH 6.8, 20% Glycerol, 0.02% Bromophenol Blue, 285mM β-MeOH) accordingly to their cell content for their normalization (either 50 or 100x OD_{600} μl of Protein Loading Buffer; i.e. for a

0.6OD_{600}, 30 or 60µl of loading buffer is added). They were subsequently boiled 10 to 15 minutes in order to break totally the cell membrane.

4.2. Protein quantification

Protein concentration was calculated using the Beer-Lambert Law, $A_{280} = \varepsilon lc$ (where A_{280} = Absorbance at 280nm, ε = Extiction coefficient, l = path lenght, c = Concentration). The protein absorbance was measured with a Nanodrop ND-1000 spectrophotometer, and the extinction coefficients were calculated from the amino acid composition using the EXPASY algorithm of denatured proteins (±5%) (Edelhoch 1967; Gill and von Hippel 1989). No difference in the concentrations were quantified when the SH3 proteins were denatured with 6M Guanidinium Chloride, therefore, regular measurements were performed without denaturing the protein. All the proteins were kept at 4ºC for short term usage, and at -80ºC for long term storage.

Alternatively, protein concentration was estimated with the Bradford (BioRad 500-0006) and BCA (Pierce 23227) standard assays, following the protocols provided by the manufacturers.

5. Crystallography

"Si la proteína no cristaliza a temperatura ambiente, no es interesante para la ciencia"

Macromolecular X-ray crystallography is a powerful tool for solution of protein structures since, provided that adequate ordered crystals are available, there is almost no limit to the complexity of the structures that can presently be determined by X-ray diffraction. It is though, a relatively young area of research, given that it was only five decades ago that Max Perutz and John Cowdery Kendrew were awarded with the Nobel Prize in Chemistry for the solution of Sperm Whale myoglobin by X-ray crystallography. Since then, the number of protein structures solved has grown exponentially with more than 5000 proteins released to the Protein Data Bank (PDB) public database in the last year (4451 already in the present year). This growth is due not only to the improvement of the technical devices, but to a better understanding of the technique and crystallization process and to the development of new data processing software. The contribution of X-ray crystallography to the structural biology is especially important, given that from the 45865 structures in the database (44052 of them being of protein structures), 85% have been solved by this technique, while Nuclear Magnetic Resonance (NMR) has contributed with >14% of them, and electron microscopy with 154 structures (data obtained from the PDB, September 2007).

The reason we use X-ray diffraction to solve protein structures is because in order to visualize objects by electromagnetic radiation, incoming radiation needs to have a wavelength comparable to the smallest features that you wish to resolve. The distance of carbon bonded atoms falls into the X-ray wavelength range in the

electromagnetic spectrum, so that by radiating a protein crystal with X-rays we obtain a diffraction pattern with information on the distribution of the electrons in the molecule, which will allow us to build a model of the protein structure in the crystal.

A single molecule in solution has not enough scattering power alone and would be masked by the scattering of water and air. Crystals arrange huge numbers of molecules in the same orientation, so that scattered waves can add up in phase and increase the signal to a measurable level. And this is the reason why we need to obtain protein crystals for the use of this technique.

5.1. Protein crystallization

"A lot of protein crystallization is dumb luck"

Bob Cudney

Crystallization of macromolecules is not trivial and it is the least understood step in the X-ray structural analysis of a protein (Drenth 1994). As mentioned previously, there is no apparent relationship between a protein's properties and the conditions under which it will crystallize.

Protein crystallization occurs when the concentration of protein in solution is greater than its limit of solubility and so the protein is in a superstaturated state. Traditional methods of crystallising inorganic molecules have been modified to be gentle enough for proteins, which are sensitive to temperature and high concentrations of organic solvents. Many methods exist to crystallise proteins, but the two most successful methods are the *microbatch* and *vapor diffusion* techniques. Concentrated solutions of the protein are mixed with various solutions, which typically consist of a buffer to control the pH of the experiment, a precipitating agent, to induce supersaturation (typically Poly ethylene glycols (PEGs), Salts such as Ammonium sulphate or organic alcohols) and other salts or additives, such as detergents or co-factors.

Both microbatch and vapor diffusion methods allow the protein solution to concentrate over time, feature that favors nucleation of the protein. Two common vapor diffusion methods for crystallisation are *hanging* and *sitting drop*. In the hanging method, a small drop of concentrated protein- and precipitant-containing solution is applied to a glass coverslip which is then inverted so as to suspend the drop above a larger reservoir of a similar solution lacking protein but containing a higher concentration of precipitant, and the chamber is then sealed (Figure IV.2). In the sitting drop method the drop is placed in a pedestal above the reservoir solution (Figure IV.3). Over time, the drop containing protein equilibrates with the larger reservoir beneath it as water in the drop leaves the drop and transfers to the reservoir, effectively increasing the precipitant concentration in the protein drop. Supersaturation of protein is favored only in some solutions that cannot be predicted and are specific for each protein construct, and it is then that *crystal nuclei* form, leading to crystal growth. Typically, hundreds or thousands of conditions are screened before a suitable condition is found that leads to a crystal of suitable quality; therefore, most of the effort for the solution of a protein structure is put into the crystallization of such protein.

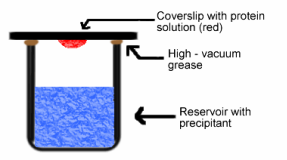


Figure IV.2. Diagram of hanging drop method. Reservoir solution (blue) usually contains buffer and precipitant. Protein solution (red) contains the same compounds, but in lower concentrations. The protein solution may also contain trace metals or ions necessary for precipitation of particular proteins (taken from {http://www.bio.davidson.edu/courses/MolBio/MolStudents/spring2003/Kogoy/protein.html}).

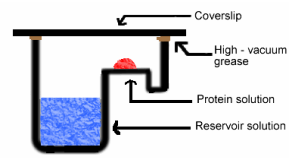


Figure IV.3. Diagram of sitting drop method. In this method, the protein drop sits on a pedestal above the reservoir solution, as opposed to hanging (taken from {http://www.bio.davidson.edu/courses/MolBio/MolStudents/spring2003/Kogoy/protein.html}).

Our first approach for initial search of chemical conditions that would deliver diffraction-quality crystals began with sitting drop vapor diffusion crystallization trials with sparse matrix screens, Polyethylene Glycol (PEG) matrices or home made screens (see Table IV.7): the Classics (Cat. #130901 & 130701), PEGs Suite (Cat. #130904), Cryos Suites (Cat. # 130903) and pHClear (Cat. #130909) suites from Qiagen; Crystal Screens I & II (Cat. # HR2-110 and HR2-112) and Index (Cat. # HR2-134) from Hampton Research; JBScreen Classic I and II (Cat. # CS-101L & CS-102L), Wizard Screens I + II (Cat. #EBS-WIZ-F) and Wizard Cryo™ I + II (EBS-CRYO-F) from Jena Bioscience.

Table IV.7. Crystallization screens used for inicial search of crystallization conditions.

Screen	Company	Cat. #
NeXtal Classics Suite	Qiagen	130901 & 130701
NeXtal PEGs Suite	Qiagen	130904
NeXtal Cryos Suite	Qiagen	130903
NeXtal pHClear Suites	Qiagen	130909
Crystal Screens I & II	Hampton Research	HR2-110 & HR2-112
Index Screen	Hampton Research	HR2-134
JBScreen Classic I and II	Jena Biosciences	CS-101L & CS-102L
Wizard Screens I + II	Jena Biosciences	EBS-WIZ-F
Wizard Cryo™ I + II	Jena Biosciences	EBS-CRYO-F
A (PEGs)	Home Made	N/A
B (Various)	Home Made	N/A
C (Salts)	Home Made	N/A

Different 96-well sitting drop crystallization plates were used from different companies for initial screenings: Innovaplate, Corning, and Costar. 24-well hanging and sitting drop plates were used for grid screening for crystal optimization.

Table IV.8. Crystallization plates used for crystal growth.

Company	Description	Cat. #
96-well plates		
Hampton Research	96x1 well sitting drop plate, Low Profile, Polystyrene, Greiner, 80 plates	HR3-093G
Hampton Research	96x3 wells sitting drop plate, hydrophobic surface, Greiner, 40 plates	HR8-149
Hampton Research	96x1 wells sitting drop plate, untreated, Greiner, 80 plates	HR3-304
Corning	CrystalEX™ 96 Well Flat Bottom Protein Crystallization Microplate, Advanced Polymer, 25/Bag, Hydrophilic Treated, Nonsterile	3785
Innovadyne	Innovaplate™ SD-2 Crystallography Plate, Package of 100	11978-1
24-well plates		
Qiagen	100 x 24-well plates each with 24 crystallization supports	132025
Hampton Research	Cryschem Plate, 24 well sitting drop, 100 plate case	HR3-160
Hampton Research	Hanging drop, 24 well VDX™ Plate, 100 plate case	HR3-140
Hampton Research	Hanging drop, 24 well VDX™ Plate with sealant, 100 plate case	HR3-170

Often, the initial screens were as much a test of protein concentration as of crystallization conditions. If approximately 30% of the conditions yielded a precipitate, we consider the protein concentration to be appropriate for further trials. If not, the concentration of the protein was adjusted accordingly. For identification of optimal initial protein concentration, Hampton Pre-Crystallization Test (Cat. #HR2-140) and Nextal EasyXtal Pre-Screen Assay were occasionally used as a guide.

The use of nanodroplet high-throughput crystallization allows for numerous conditions to be sampled while keeping the requirements of protein quantities to a minimum (Stevens 2000) and therefore, reduces the burden on protein production. We used a medium high-throughput system in which either 0.5+0.5μl or 0.8+0.8μl protein+reservoir solution drops were set manually with a multi-channel pipette or 100+100ng protein+reservoir solution drops were set with the Cartesian MycroSys synQUAD liquid handling robot in 96-well crystallization plates. Plates were incubated at 22ºC or 4ºC temperature monitored rooms.

Drops for crystallization complexes of SH3 domains with synthetic peptides were set with a protein-peptide mixture of 1.5 to 3 times peptide molar excess. The peptides were diluted in 100mM Tris pH 8 and the mixture was incubated 10 to 15 minutes at RT for proper complex formation. Prior to drop setting, the sample was centrifuged at 14000xg for 10-15 minutes to bring down any particle or aggregate present in solution.

All Cbl derived peptides used for crystallization and binding assays were synthesized from C to N-terminus by Sigma-Genosys using Fmoc chemistry and solid support resin, with >90% purity guaranteed.

Peptide sequences are summarized in the following table (Table IV.9). Mutations of native sequences are shown in red, and residues from the CIN85/CMS binding motif are aligned and highlighted in bold letters. An N-terminal tyrosine was added to some peptides for accurate quantification of peptide concentration for the ITC binding assays.

Table IV.9. Polyproline peptides used for co-crystalization and binding studies.

Peptide	Sequence CIN85 binding motif RxPx PxPR
c-Cbl$_{811-834}$	TNVTEGSQVPE**R**P**P**K**P**F**PR**RINSE
c-Cbl$_{816-834}$	GSQVPE**R**P**P**K**P**F**PR**RINSE
c-Cbl$_{820-830}$	PE**R**P**P**K**P**F**PR**R
Clb-b$_{893-914}$	PSCSDGSQAPA**R**P**P**K**P**R**PR**RTA
Cbl-b$_{898-914}$	GSQAPA**R**P**P**K**P**R**PR**RTA
Cbl-b$_{902-912}$	PA**R**P**P**K**P**R**PR**R
Cbl-b$_{902-912}$R911A	PA**R**P**P**K**P**R**A**RR
YCbl-b$_{903-912}$	Y**A**R**P**P**K**P**R**P**R**R
YCbl-b$_{903-912}$R904A	Y**A**A**P**P**K**P**R**P**R**R
YCbl-b$_{903-914}$	Y**A**R**P**P**K**P**R**P**R**RTA
CD2$_{324-333}$	KGP**P**L**P**R**P**RV
YCD2$_{325-335}$	YGP**P**L**P**R**P**RVQP

5.2. Crystal cryo-cooling

Protein crystals, as well as living tissues, undergo radiation damage if exposed to X-rays. Radicals are formed in the crystals upon exposure to high-energy X-ray photons. These radicals lead to subsequent chemical reactions that gradually destroy the crystalline order and as a consequence the information that can be obtained from a single crystal is limited. Radiation damage is proportional to dose and might be structurally specific. It has been calculated that each absorbed 8keV photon disrupts ~70 molecules and disorders another 90. Colin Nave and Elspeth F. Garman reported that diffraction at 100ºK usually prolongs the crystal lifetime by a factor of ~70 (Nave and Garman 2005), normally long enough for a complete dataset to be collected from a single crystal.

Among the consequences of radiation damage are the decrease of diffraction intensity and resolution, the increase in the unit cell volume, and site-specific damage. The latter occurs in a well-defined order, starting with the breakage of disulphide bonds, followed by decarboxylation of aspartates, glutamates and the C-terminus, and then loss of the hydroxyl group from tyrosines. Active sites and metal centres appear to be particularly vulnerable and, thus, they must be inspected carefully when assigning structural or functional significance (Nave and Garman 2005; Ravelli and Garman 2006).

This is why nowadays X-ray data collection from biomacromolecular crystals is usually performed at cryogenic temperatures. However, crystal cryo-cooling needs to be carefully undertaken. Slow cooling allows water inside and surrounding the crystal to form hexagonal ice, disrupting the order of the protein lattice and creating ice rings that obscure its diffraction. The protein conformation might even be altered upon formation of ice within the crystal. To avoid water contained in the solvent to

transform into ice, a rapid or `flash' cooling of the crystal is usually performed so that any water content is transformed into an amorphous (vitreous) form and its redistribution within the crystal is prevented. This procedure generally improves the diffraction results (Kriminski, Kazmierczak et al. 2003).

Transforming pure water to amorphous ice requires cooling from room temperature to below water glass transition temperature in less than 10^{-4}s. The use of penetrating cryoprotectants such as glycerol or ethylene glycol and non-penetrating cryoprotectants such as large molecular weight polyethylene glycols (PEGs) (commonly present in the crystallization mother liquor as precipitants) can increase cooling times required for vitrification of internal and external solvent. Protein itself can be an excellent cryoprotectant at specific concentrations usually found in protein crystals (Kriminski, Kazmierczak et al. 2003).

We used glycerol or ethylene glycol-containing cryoprotectant solutions to wash the protein crystals and ensure collection of single crystals with no precipitate content. When components of the crystallization solution itself acted as cryoprotectants, crystals were washed in reservoir solution.

Protein crystals are usually flash-cooled by inserting a crystal into a nitrogen or helium cold gas stream, or by plunging into liquid nitrogen at 77ºK or liquid propane at 86ºK. We flash-cooled the crystals of this study under a nitrogen gas source using an Oxford Cryosystem device.

5.3. Crystal data collection

Once the crystal is ready for the experiment, it can be exposed to an X-ray beam. Structural information is obtained from crystal diffraction.

X-rays are scattered in all directions upon interaction with the electrons contained in the protein crystal. Depending on the distance traveled by the scattered rays, their relative phases will differ or not. When scattered X-rays are in phase, they add up with a resulting amplitude that is the sum of the individual amplitudes. On the other hand, when they are out of phase, their amplitudes cancel out.

When the angle of incidence is the same as the angle of reflection, incident and scattered light rays are in phase. Such condition in crystallography is true only for Bragg planes. The crystal lattice defines the distance that separates these Bragg planes, and only electrons situated on this theoretical plane contribute to the scattering of the incident X-ray. Given that X-rays are scattered in all directions, there is an infinite number of Bragg planes (see Figure IV.4).

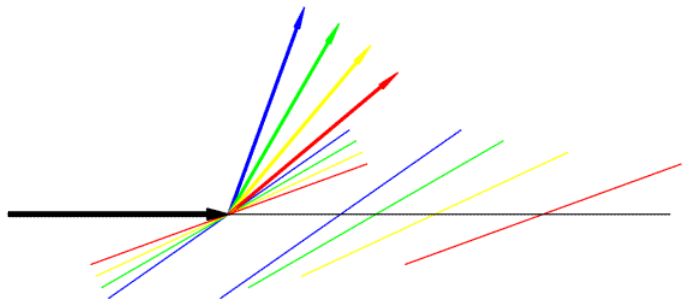


Figure IV.4. Representation of pairs of Bragg planes. The large black arrow represents the incoming radiation. Coloured arrows represent different diffracted rays, and a pair of Bragg planes is shown in the same colour (taken from the Macromolecular Crystallography course of 1999-2000, University of Cambridge and the MRC-LMB: http://images.google.es/imgres?imgurl=http://www-structmed.cimr.cam.ac.uk/Course/Basic_diffraction/Ewald_planes.gif&imgrefurl=http://www-structmed.cimr.cam.ac.uk/Course/Basic_diffraction/Diffraction.html&h=259&w=600&sz=13&hl=es&start=9&um=1&tbnid=bDDovWxsvq5VTM:&tbnh=58&tbnw=135&prev=/images%3Fq%3Dewald%2Bsphere%26svnum%3D10%26um%3D1%26hl%3Des%26client%3Dfirefox-a%26rls%3Dorg.mozilla:es-ES:official%26sa%3DN).

In order for the unit cells to diffract in phase, the Bragg planes must pass through the same points in all the unit cells in the crystal. The easiest case to imagine is when the planes are separated by one unit cell edge, or an integral fraction of a unit cell edge. On the other hand, if the planes divide the unit cell edge by a non-integral number, the different unit cells will all diffract out of phase and the waves will cancel out.

In order to help visualise which Bragg planes are in the correct orientation to diffract we can construct the Ewald sphere of radius 1/λ. In such sphere, only scattered rays from the Bragg planes that cross the Ewald sphere will be in phase (see Figure IV.5). The diffraction pattern of a crystal thus, can therefore be used to extract information about its symmetry and arrangement of its electrons.

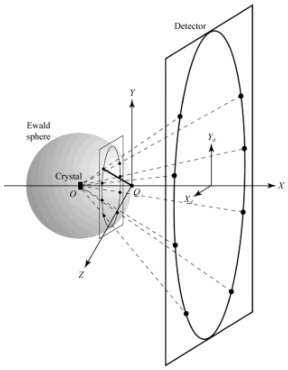


Figure IV.5. The Ewald sphere construction. The X-ray beam is along the X axis and the Z axis is the rotation axis. The origin of the reciprocal lattice lies at the point that the X-ray beam exits the Ewald sphere (Q) and the crystal is located at the centre of the sphere (O). The crystal is oriented so that the a axis lies along the X axis. The reciprocal-lattice plane h = 1 is shown. Each diffraction spot on the detector can be mapped back to the equivalent scattering vector in reciprocal space. One such scattering vector is shown as a bold line from the reciprocal-lattice origin (Q) to the surface of the Ewald sphere. Taken from (Leslie 2006).

A common experimental setting used to collect the crystal diffraction pattern is such in which the crystal is positioned to be directly exposed to the incident X-ray beam, and a detector is located in the other side of the incident beam to record the diffraction pattern. In order to collect all the atomic information contained in the

crystal, either the detector needs to be rotated around the Ewald sphere or the crystal needs to be rotated around itself. For technical ease, it is usually the latter that is used.

We collected the diffraction images of CIN85-SH3C$_{270-328}$His$_6$ and CIN85-SH3A$_{2-58}$His$_6$ crystals on a Bruker F530 rotating anode X-ray source at 110ºK, 1.54179ÅA, with a Mar345 image plate, which is a photosensitive detector. In the case of the CIN85-SH3A$_{2-58}$:Cbl-b$_{902-912}$ crystal a Bruker-Nonius generator was used for collection at 100ºK, and a Kappa2000 CCD detector.

Crystals from CIN85-SH3A$_{2-58}$:CD2$_{324-333}$ and CMS-SH3A$_{1-62}$:YCbl-b$_{903-914}$ complexes were collected under the more powerful synchrotron radiation source at the European Synchrotron Radiation Facility (ESRF) beamlines ID23-1 and ID14-1, respectively, with CCD detectors.

5.3.1. Data reduction

Once the data is collected, it needs to be processed to obtain the unit cell parameters and the orientation of the crystal, which may indicate the symmetry of the crystal. With that information, the images can be integrated to predict the positions of the Bragg reflections on each image and obtain an estimate of the intensity of each reflection and its uncertainty (Leslie 2006). There are several programs available to carry out such data reduction.

We used Denzo (Otwinowski 1997) program for the processing of CIN85-SH3A$_{2-58}$His$_6$ and CIN85-SH3A$_{2-58\ tga}$:Cbl-b$_{902-912}$ data, and the program Mosflm (Collaborative Computational Project 1994) for all the other data sets.

There are several physical factors that affect the measured diffraction intensities, such as variations in the crystal rotation rate, variation in incident-beam intensity (in synchrotron radiation), radiation damage or errors from the detector. Therefore, the diffraction data need to be scaled to put them on a consistent scale. After such scaling, intensities can be analysed to establish the real resolution of the data set, to detect bad regions, to analyse radiation damage and to assess the overall quality of the data set (Evans 2006).

In our study, hkl-Scalepack (Otwinowski 1997) was used to scale CIN85-SH3A$_{2-58}$His$_6$ and CIN85-SH3A$_{2-58\ tga}$:Cbl-b$_{902-912}$ data, and Scala (Collaborative Computational Project 1994) to scale data that was reduced with Mosflm.

5.3.2. Structure solution by Molecular Replacement (MR)

The diffraction pattern we measure in the diffraction experiment is essentially the result of a count of the number of X-ray photons in each spot. These spots depend directly on the phase of the scattered wave, since, as already

mentioned, amplitudes of scattered waves that are in phase will add up, and the amplitude of scattered waves that are out of phase will cancel out. However, there is no practical way of measuring the relative phase angles for the different diffracted spots experimentally. The number of photons gives the intensity, which can be turned into amplitudes. But the phase has been lost.

The diffraction pattern is related to the object diffracting the waves through a mathematical operation, the Fourier transform. Diffraction patterns gives us information of the "reciprocal space", which, calcultating the inverse Fourier transform, gives information of the "real space". To compute the inverse Fourier transform we need the amplitude and the phase of the diffracted waves (see Figure IV.6).

Figure IV.6. A particularly dramatic illustration of the importance of phase. On the top are photographs of Jerome Karle (left) and Herb Hauptman (right), who won the Nobel Prize for their work on solving the phase problem for small molecule crystals. The photographs are treated as density maps and their Fourier transforms are calculated to get amplitudes and phases. When the phases from the picture of Hauptman with the amplitudes from the picture of Karle are combined, we get the picture on the bottom left. The bottom right picture combines the phases of Karle with the amplitudes of Hauptman. Clearly the phases are dominating what we see (http://www-structmed.cimr.cam.ac.uk/Course/Fourier/Fourier.html#phase).

One way to solve the phase problem is to possess an atomic model, from which estimates of the phases can be computed. Such a model can be obtained when the structure of a related protein is known, or even the same protein from a different crystal form. But to build up an atomic model of the new crystal form, it is necessary to work out how the model should be oriented and positioned in the new cell. Molecular replacement is the technique used to solve that problem.

Molecular replacement can be used to solve a structure when a good model is available for a reasonably large fraction of the structure in the crystal. As the number of solved structures in the database increases, this method will be useful for a larger fraction of new structures. Of course, it is also useful for binding studies or studies of complexes formed from previously-determined proteins.

The level of resemblance of two protein structures usually correlates well with the level of sequence identity, which means that it can be speculated whether or not molecular replacement will succeed before even trying it. As a rule of thumb, molecular replacement will probably be moderately straightforward if the model is reasonably complete and shares at least 40% sequence identity with the unknown structure. It becomes progressively more difficult as the model becomes less complete or shares less sequence identity.

This technique was particularly appropriate for the solution of the SH3 structures, since all SH3 domains have essentially the same overall structure. For the first SH3 domain that was solved in this project (CIN85-SH3C), SEM5 C-terminal SH3 domain (PDB code 1SEM) was used as a search model, since they have a sequence identity of 47%. For posterior structures, CIN85 and CMS SH3 domains were chosen as search models. Molecular replacement was performed using the programs AmoRe (Navaza 1994) and Molrep (Collaborative Computational Project 1994).

In the case of CIN85-SH3C and CIN85-SH3A:CD2, after the theoretical phases were assigned automatic model building was carried out using the program ARP/wARP (Zwart, Langer et al. 2004).

5.3.3. Model refinement

Once the structure solution is found, a model can be generated either by automatic model building or, alternatively, model building can start directly from the MR solution. The parameters of this model need to be changed to be optimized so that it agrees with the observed data. Refinement of a structure model consists on the optimization of a function of a set of observations.

In refinement programs, several properties are taken into account, such as crystal symmetry, stereochemical properties of the atoms in the molecule, the atoms' vibrations about their central positions (which estimates are indicated by the B factors), or the internal energy of the molecule, among others.

A diffraction data set at 2Å resolution of a medium-sized protein with about 2500 atoms contains around 22000 reflections. This means that in order to be able to handle the data, the number of parameters to be refined need to be minimized.

There are a number of programs available for structure model refinement. Each of them makes their own set of assumptions, with certain distribution of errors and different consequences of such errors. Therefore, one may work better than another in a particular problem.

Two main statistical methods are used for model refinement: least squares and maximum likelihood. The latter does not need to make the assumptions required for the least squares method (which have been shown to be incorrect in many refinement problems), and it is more broadly used nowadays (Tronrud 2004).

For the CIN85-SH3A$_{2\text{-}58}$His$_6$ model, automatic refinement was carried out with CNS software (Brunger, Adams et al. 1998), combined with manual model building in Xfit (McRee 1999).

All the other models were refined by cycles of positional and restrained individual B-value automatic refinement with Refmac5 software (Collaborative Computational Project 1994) and manual model building using Coot (Emsley and Cowtan 2004) software. In the case of CIN85-SH3A:CD2 complex, final refinement was carried out with Shelxl software (Sheldrick, Schneider et al. 1997).

5.4. Protein characterization

5.4.1. Fluorescence Spectroscopy

Proteins are the only biomolecules that have an intrinsic fluorescence useful and measurable, in contrast to lipids, membranes and saccarides that are in principle non fluorescent, or DNA, which fluorescence is too weak to be useful. The three aromatic residues (phenylalanine, tyrosine and tryptophan) contribute to the protein ultraviolet fluorescence, and, to our advantage, they are uncommon in protein sequences. Particularly, tryptophan absorbs at the longest wavelength and displays the largest extinction coefficient so that it dominates the emission of proteins. Tryptophan is especially sensitive to its local environment, and thus, changes in its emission spectrum in response to conformational transitions, substrate binding or denaturation, among others, can be observed. All of these processes can affect the local environment of the indol ring of the tryptophan. In addition, tryptophan has a tendency to donate electrons from the indol group when it is in an excited state, which gives it a unique sensitivity to *collisional quenching*[5]. The decrease in fluorescence, or *quenching*, can be caused by either an external molecule, such as a protein substrate, or by nearby groups in the protein (Lakowicz 1999).

To measure the change in fluorescence emission upon peptide addition, a 1-2μM protein solution and a 1.5-2.5mM peptide solution were prepared in the Superdex elution buffer. The lyophilized peptides were weighted in an analytical balance and corresponding buffer volume was added for their solubilization. Protein concentration was calculated by measurement of absorbance at 280nm using the Beer-Lambert Law as described in section IV.4.2. Samples were mixed and temperate in a Peltier stirring block to 20ºC. Excitation and emission slits were open 2 and 4mm, respectively and the photomultilplier set to 1000V.

5ml cuvettes were filled with 2.5ml of protein solution, and samples were excited at 290nm to avoid contribution of the emission of phenylalanine and tyrosine. Emission spectra were measured from 310 to 370nm in 1nm steps. Between each spectrum measurement, 0.3μl of peptide was pippeted in manually, and incubated under continuous stirring at 500rpm for 15 minutes at 20ºC. A tryptophan bleaching control was taken by measuring protein against buffer emission spectra and the same was done with buffer against buffer for dilution correction.

Buffer measurement was substracted to the raw data to correct for dilution. Subsequently, an average of the five maximum points was calculated (335-340nm) and tryptophan bleaching correction was applied (Di Stasio, Bizzarri et al. 2004). For the latter, dilution correction and average of the peak was calculated for control data, time-spots vs. relative fluorescence (normalized) was plotted to calculate the parabolic equation which was then applied to each point in the spectra. Every point

[5] The decrease in intensity of fluorescence is referred as *quenching*. *Collisional quenching* occurs when the excited-state fluorophore is deactivated upon contact with some other molecule in solution (which can be referred as the *quencher*).

of the corrected data was divided by the calculated tryptophan bleaching contribution and a fit curve could be calculated with Origin software applying the single-site binding model equation ($F_{rel}=(1+K[L]F_{ML})/(1+K[L])$, where F_{rel}= Relative Fluorescence, normalized to F_0); K= association constant, for which the published data (14μM K_D) was used as a temptative value; [L]= ligand concentration, F_{ML}= Fluorescence of the complex) (Eftink 1997).

5.4.2. Isothermal Titration Calorimetry (ITC)

Almost all physical or chemical processes have an associated heat effect. In most biological binding events, heat is absorbed or generated, and this can be used as the basis for a number of analytical techniques as well as for the determination of absolute thermodynamic quantities for the study of biomolecular interactions. Microcalorimetry is a good biophysical method for non-invasive, non-destructive analysis of biomolecular stability and interactions, since no tagging or immobilization of binding components is needed (http://www.microcal.com/index.php?id=312).

Titration calorimetry is the most direct method to measure the heat change on formation of a complex at constant temperature. In a typical Isothermal Titration Calorimetry experiment a ligand is titrated into a solution of protein (or *vice versa*), and the association constant, K_A, and the stoichiometry, n, of the complex are obtained (Jelesarov and Bosshard 1999). After each addition of a small aliquot of ligand, the heat released or absorbed in the sample cell is measured with respect to a reference cell filled with buffer so that the quantity of heat absorbed or released is in direct proportion to the amount of binding occurring. When the system reaches saturation, the signal diminishes and only heats of dilution are observed.

The setup of our experiment in a VP ITC Calorimeter from MicroCal consisted on titration of a peptide solution in the protein buffer concentrated 10 times more than the protein. Protein and peptide solutions were degassed and taken to 18ºC temperature in a Thermovac apparatus from Microcal. Protein concentration used varied from 1 to 2 μM, and peptide was titrated in 2 and 5μl volumes in 30 second intervals. Heat measurements were taken at 18ºC.

Control measurements were taken with titrations of buffer into buffer and buffer into protein, so that heat emission of solvent dilution could be substracted to the raw data. Data were fit by least-squares procedures assuming a one-site binding model using Microcal Origin software version 7.0 (function *itcfun32.2*).

5.4.3. Circular Dichroism (CD)

Circular dichroism measures the wavelength dependence of the differential absorption of right and left circularly polarized light. The right and left components of a polarized beam of light interact differently with the chiral centers of an optically active chromophore. This interaction results in a decrease of the velocity of wave propagation and its absorption, which depends on the wavelength. This fenomena results in a difference in phase of each component of the polarized light so that the sum of the vectors of the two components rotates following a helical path.

Measurement of the differences in reflected light gives an idea of the chirality of the molecule and its structure.

Because of the multiple conformations adopted by the peptide bond (the principal element detected by CD), the CD spectrum is the result of an average of the various conformation parameters. Therefore, the precise structure of a protein cannot be constructed from its CD spectrum. Nevertheless, this technique is broadly used for comparison with spectrums of proteins with already known structures and gives an idea of the structural features of the molecules (Freifelder 1982).

The different secondary structure elements display characteristic spectra, so that we can predict which of them prevails: antiparallel β-sheets, turns or helices.

Measurements were taken for individual His-tagged SH3 domains from CIN85 in a Jasco J810 apparatus. The system was purged with liquid nitrogen to evacuate all the oxygen, and gas speed were set to 15-20 L/min.

The concentration of the protein was adjusted so that the signal was lower than 800AU (~21μM). 2mm cuvettes were used with 0.4ml of protein solution. 20 cycle measurements were taken at 20ºC.

A temperature denaturing curve was collected at the maximum peak, at 1ºC/min. Subtraction of the buffer baseline was not needed since the protein had been diluted 300 to 1000 in water and the buffer spectra showed values very close to zero.

5.4.4. Chemical shifts in monodimensional Nuclear Magnetic Resonance (NMR)

As seen for circular dichroism spectrometry, 1-D nuclear magnetic resonance (NMR) can be used to identify secondary structure elements of a protein. Comparison of empirically measured chemical shifts of a protein in monodimensional NMR spectra with those of known-three dimensional structures allows the assignment of secondary structure elements such as β-sheets, α-helix or random coils.

In a molecule, each nucleous of the same type will have a resonance frequency that depends on the chemical group in which it resides, which means that shifts in chemical resonance (*chemical shifts*) of backbone atoms in proteins are extremely sensitive to local conformation. Advanteously, homologous proteins show similar patterns of secondary chemical shifts that can be used to assign secondary structure features to proteins of unknown structure.

Purified CIN85 SH3 samples were diluted to 0.1mM in PBS buffer and dialysed against PBS to get rid of any Hepes traces from the Superdex buffer. Deuterium was added to a 10% final concentration.

One dimensional NMR experiments were performed at 25ºC in a 600MHz Bruker Avancer spectrometer and analized by Dr. Daniel Padró and Dr. Francisco J. Blanco from the NMR group at the CNIO. XWinNMR version 3.5 (Bruker) was the software used for the acquisition, processing and analysis of all spectra.

RESULTS

“Be aware that negative results can be just as important as positive results”

E. R. Schulman

“Un resultado negativo es un resultado”

J. Bravo

V. Results

1. CIN85

1.1. Individual CIN85 SH3 domains

In our attempt to study the interactions that take place in the signalosomes required for EGFR endocytosis, we successfully cloned and purified several CIN85 and c-Cbl fragments, in addition to wild type human Ubiquitin. Each clone was designed with GST and hexahistidine affinity tags at different positions, as well as un-tagged: full length CIN85, $CIN85_{1-607}$, $CIN85_{1-600}$, CIN85-SH3ABC, CIN85-SH3AB, CIN85-SH3BC, $c\text{-}Cbl_{808-906}$, $c\text{-}Cbl_{816-906}$, $c\text{-}Cbl_{816-895}$, $c\text{-}Cbl_{820-830}$ and full length Ubiquitin. Expression of all of them was endeavored in different systems and they were purified to be combined to form different multimeric complexes which interactions were confirmed by different methods; nevertheless, crystals obtained in this approach, if any, were not suitable for structural determination.

It is widely known in the field that proteins with multiple domains, such as CIN85, are difficult targets for structural studies for several reasons: "*First, multi-domain proteins are difficult to express in E. coli",* which is a low cost, fast and easy to handle expression system. *"Second, these proteins exhibit conformational heterogeneity, which decreases the probability of crystallization*" (Edwards, Arrowsmith et al. 2000).

SH3 domains are generally very stable and globular domains, thus, as an alternative and in order to simplify the crystallization problem and focus on the study of the direct interaction between CIN85 and Cbl, we expressed and purified the SH3 domains individually.

Characterization of the CIN85 SH3 domains both individually and complexed with synthetic peptides derived from the Cbl-b and c-Cbl CIN85 binding sequences should provide specific information regarding the differences among the CIN85 SH3 domains. It should also offer some clues regarding SH3 differential or redundant functionality within the CIN85 molecule, a recurrent issue expossed in the literature. Their characterization would allow us to compare binding of other SH3 domains to different polyproline motives, and thus, be able to elucidate the relevance of the described novel binding motif (Kowanetz, Szymkiewicz et al. 2003; Kurakin, Wu et al. 2003) in comparison to the typical xPxxP motif and other already reported atypical SH3 binding motifs (Lewitzky, Harkiolaki et al. 2004);(Mongiovi, Romano et al. 1999);(Berry, Nash et al. 2002);(Tong, Drees et al. 2002).

1.1.1. Design of SH3 constructs and the Cbl peptides

SH3 domains provided by I. Dikic were designed so that 22 residues reminiscent from the vector sequence, which represent a ~37% of the construct

sequence length, were expressed (see Figure V.1). This fact should not represent a problem for interaction and cell biology studies, yet it might certainly be an obstacle for protein crystallization. Crystallisation trials of these constructs did not yield any crystals suitable for any further structural characterization so that we proceeded to the design of new CIN85 individual SH3 constructs.

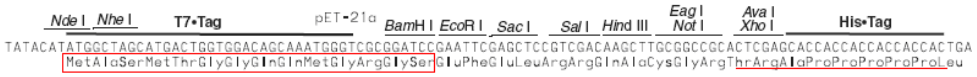


Figure V.1 Multiple cloning site (MCS) of pET-21a from Novagen. Nucleotide sequence of the MCS is transcribed in the bottom line. On the upper part, sequence sections corresponding to the differentent restriction enzyme (RE) sites and expression tags are indicated. On the lower part, the amino acids corresponding to the DNA translation are annotated. The hexahistidine tag is expressed when in frame with the Xho I site. Vector aminoacids expressed in the clones provided by Ivan Dikic are indicated in red.

As an alternative to increase the chances of obtaining suitable crystals, we proceeded to the recloning of the SH3 constructs, mainly to avoid the expression of extra residues coded into the vector sequence. Expressed residues foreing to the protein sequence do not provide any relevant structuctural information inherent to the protein; however, they can provide the protein constructs with flexible fragments. Flexibility is an attribute that plays against crystal formation, given that in a crystal, the molecules must pack together in a regular, periodic, and continuous manner. Flexible fragments might thus grant conformation variability to the protein sample, and thus, do not allow formation of the periodic arrangement that builds up a protein crystal. As a matter of fact, it has been found that removal of mobile elements from a protein by proteolysis or by molecular biology methods, allows the prompt crystallization of proteins that had not been possible to crystallize by all means prior to exclusion of such elements (McPherson 1999).

We compared the CIN85 SH3 domains with the sequence of a highly identical SH3 domain of which the structure was known. This comparison would allow us to evaluate whether the extension of the domains was appropriate and hence allow us to limit it to the residues strictly necessary to fold into an SH3 domain. A BLAST search in the Protein Data Bank (PDB) database with the CIN85 SH3 domains was performed, and we found that, among the SH3 domains with a known structure, the C-terminal domain of SEM5 (1SEM; *SEx Muscle abnormal family member*, GRB2 worm homolog -*Caenorhabditis elegans*-) was the one with the highest sequence identity (43%) with an expected value (E) of 0.47. Such relatively high E-value is due to the short sequence length. Comparison of the aligned sequences by ClustalW and analysis of the structure of 1SEM helped us design new constructs with an extension similar to the SEM5 SH3, avoiding flexible extremes (Figure V.2).

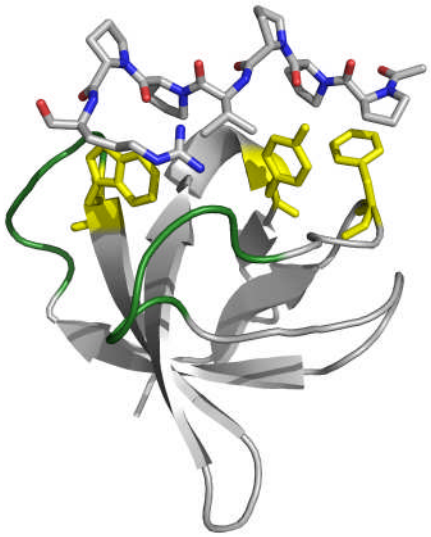


Figure V.2. Overall structure of the C-terminal SH3 domain of SEM5 bound to a polyproline rich Sos derived peptide (PDB code 1SEM). Cartoon representation of the β-barrel structure of the C-terminal SH3 domain of SEM and a stick representation of the peptide with carbons in white, nitrogens in blue and oxygens in red. Conserved aromatic residues from the binding core are highlighted in yellow, and the RT and n-Src loops that form the acidic pocket are colored in green (Figure modified from (Zarrinpar, Bhattacharyya et al. 2003)).

We cloned the new sequences into the pET-21a vector using Nde I and Xho I restriction sites that would allow the expression of a minimum number of residues foreign to the protein (Nde I codes for the initial methionine, and Xho I allows expression of only two extra residues apart from the hexahistidine tag, see Figure V.1), residues that would not provide any structural information intrinsic to the protein. In the case of CIN85-SH3C, the DNA could not be cloned into the pET-21a vector with Nde I, since such restriction site is present into the natural CIN85-SH3C cDNA sequence. Therefore, the amplicon of CIN85-SH3C was flanked by Nhe I and Xho I (this cloning causes the expression of only three extra residues in the N-terminal end, plus eight residues in the C-terminus).

In the structure of 1SEM, we observed that only the first three amino acids had a flexible character, since the first β-sheet was stabilized by a hydrogen bond established between phenylalanine 158 and the last expressed residue, asparagine 212. To be in the safe side, we limited the extension of the new designed constructs to be one residue longer in each terminus, according to the sequence alignment to SEM-$SH3_{C\text{-}term}$ (see Figure V.3). Therefore, new residue boundaries were set to residues 2 to 58 in SH3A, 101 to 157 in SH3B and residues 270 to 328 in SH3C.

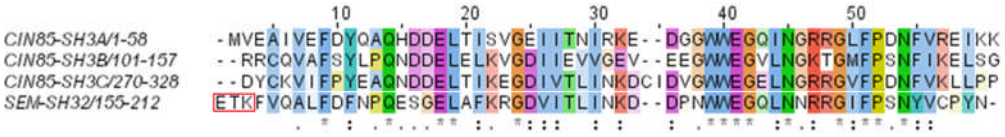


Figure V.3. Alignment of the newly designed SH3 construct sequences from CIN85 and the C-terminal SH3 of SEM5. The residues disordered in the SEM SH3 structure are indicated with a red box.

The sequence alignment of the new constructs with SEM5 $SH3_{C\text{-}term}$ individually shows an identity of around 41% with a ~22% of strong similarity and ~15% of weakly similar residues (43% identity with SH3A, 36% with SH3B and 45% with SH3C, specifically, according to the program Clustal) (see Figure V.3).

Another observation we made from the published structure of the SEM5 SH3 domain in complex with the Sos derived polyproline peptide, is that only seven of the nine amino acids in the peptide were observed in the electron density, which seemed to be the only ones contributing to the interaction. We initially had been using for crystallization trials Cbl peptides that included the sequence of those peptides previously shown to interact *in vivo* with CIN85 (Kowanetz, Szymkiewicz et al. 2003). Those extra amino acids might indeed hamper the crystallization of the complexes by providing flexibility to the peptide structure. Therefore, we decided to design new peptides with short extensions, decreasing the number of residues from 22 and 23 (in Cbl-$b_{893-914}$ and c-$Cbl_{811-834}$, respectively) to 11 (see Table IV.9 in Materials and Methods).

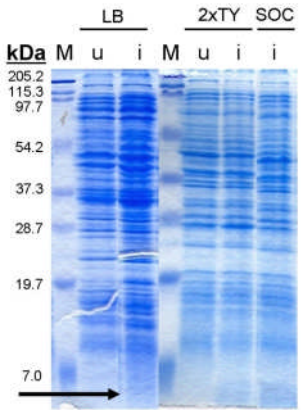
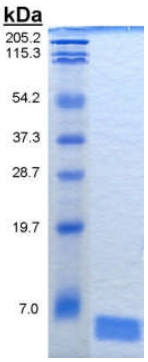


Figure V.4. Overexpression of CIN85-SH3B$_{101-157}$His$_6$ in *E. coli* strains at 37ºC. 17% SDS-PAGE gels of total cells, lanes 1 and 4: molecular weigth markers; lanes 2 and 5: uninduced (u); lanes 3, 6 and 7: induced (i); lane 3: overexpression of SH3B in BL21(DE3) with TB broth; lane 6: overexpression of SH3B in Rosetta(DE3)pLysS with 2xTY broth; lane 7: overexpression of SH3B in Rosetta(DE3)pLysS with SOC medium. The gel was stained with Coomassie Brilliant Blue.

We first pursuited the expression and purification of the new construct of SH3B. We achieved adequate overexpression of CIN85-SH3B$_{101-157}$His$_6$ in Rosseta(DE3)pLysS and BL21(DE3) *E.* coli strains with SOC, 2xTY and LB media at 37ºC (Figure V.4). Imidazole was included in low concentration in the lysis buffer to avoid competition of contaminants in binding to the nickel charged sepharose resin. An Imidazole gradient was performed in the elution segment of the affinity chromatography to selectively elute the His-tagged protein. In the final size exclusion step, we obtained a pure protein sample of an apparent molecular weight of 9.2kD, according to the gel filtration column calibration, which is close to the sequence calculated weight of 7.6kD of a monomeric protein (Figure V.5).

Figure V.5. Pure CIN85-SH3B$_{101-157}$His$_6$. Analysis of the elution peak in an SDS-PAGE shows high purity of the eluted sample. It can be observed that low molecular weight proteins run as broader bands in high percentage polyacrylamide gels.



The acquisition of suitable crystals for structural studies requires not only a chemically highly pure sample but also a structurally homogeneous protein. The SH3B protein sample was analyzed by mass spectrometry. The homogeneity of the sample was confirmed by a spectrum with a single peak of a molecular weight of 7.6kD that differed from the theoretical one in only 5Da.

In order to analyze the folding state of the protein we also performed some circular dicroism (CD) studies and monodimensional nuclear magnetic resonance (NMR) experiments. Figure V.6 shows the far UV spectrum of CIN85-SH3B$_{101-257}$His$_6$, in which we observe negative peaks around 206 and 228nm (Figure V.6A). Compared to other SH3 domain CD spectra in the literature (Figure V.6B and C), the CIN85-SH3B spectrum displays a typical pattern for SH3 domains. Additionally, the experiment was repeated after thermal denaturation to confirm total recovery of the folding state. This indicates that the thermal denaturation of the SH3 domain is totally reversible (Figure V.6A).

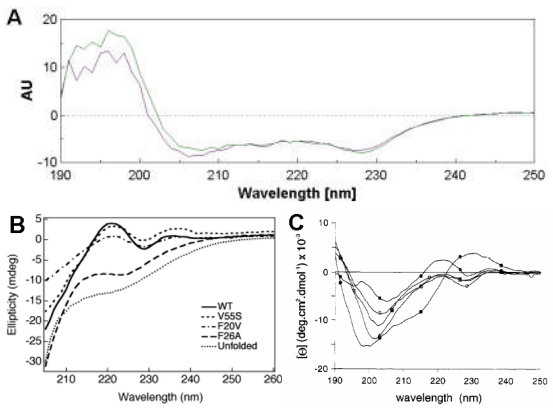


Figure V.6. Circular Dichroism (CD) spectra of CIN85-SH3B$_{101-257}$His$_6$ and other SH3 domains. (A) The green spectrum corresponds to the measurement taken previous to thermal denaturation, while the purple one represents the measurement after thermal denaturation. Both show negative peaks near 206 and 228nm. **(B, C)** Circular Dichroism (CD) spectra of **(B)** Fyn-SH3 wild type and mutants of the core domain and **(C)** SH3 domains from Abl, Fyn, c-Src, n-Src and Spectrin, studied by Di Nardo *et al.* (Di Nardo, Larson et al. 2003) and Viguera *et al*. (Viguera, Arrondo et al. 1994), respectively. The far UV spectrum of all SH3 shown, except for Abl, show similar characteristics as the one measured for CIN85 SH3B domain, with minimums around 206 and 228nm, and maximum around 222nm.

The thermal denaturation of CIN85-SH3B$_{101-257}$His$_6$ yielded a sigmoidal transition at a wavelength of 206nm. A cooperative process of denaturation occurred between 40 and 60ºC, with a melting temperature of approximately 50ºC (see Figure V.7).

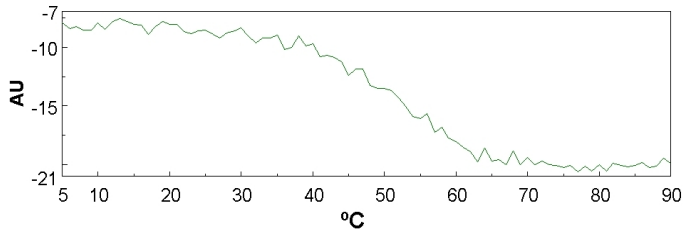


Figure V.7. Thermal stability curve of CIN85-SH3B$_{101-257}$His$_6$ monitored by far-UV CD. Mean residue weight ellipticity at 206nm, with normal two folding state behavior with a melting temperature of around 50ºC.

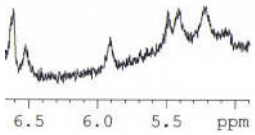


Figure V.8. Detail of the Monodimesional Nuclear Magnetic Resonance (NMR) spectra of the CIN85-SH3B$_{101-157}$His$_6$ protein construct between 6 and 5.5 ppm. Peaks in this region indicate the presence of β-sheet folding structure.

In the NMR experiment we observed chemical shifts between 6 and 5.5ppm characteristic of antiparallel β-sheet pattern in the NMR monodimensional spectrum of CIN85-SH3B$_{101-157}$His$_6$ (Figure V.8), which would correlate with the correct folding of the protein (Viguera, Arrondo et al. 1994; Viguera, Martinez et al. 1994). The

SH3B was tested for Cbl-b binding by comparison of monodimensional spectra of samples of SH3 alone, and with the peptide; however, only subtle changes were observed and results were inconclusive.

The protein sample was concentrated to 7.9mM (60mg/ml) and Wizard HT screen was set at 22ºC with drops of 0.8+0.8µl (protein+reservoir solution). Crystallization of the complex was attempted as well by sparse matrix screening of a mixture of the CIN85-SH3B protein and the long Cbl-b peptide (Clb-$b_{893-914}$) in a 1:1.5 ratio and at a calculated final protein concentration of 6.6mM (50mg/ml). We did not obtain any crystals in these conditions. However, we did not succeed in the acquisition of crystals suitable for structural determination.

1.1.2. Purification, crystallization and structure of CIN85-SH3C$_{270-328}$His$_6$

As mentioned before, CIN85 SH3 domains have been shown to associate with several intracellular signaling proteins containing different PxxxPR motifs. Interestingly, variations in these motifs can also lead to modulation of CIN85-SH3 domain binding affinities, as reported in various studies. Several authors have claimed to observe overlapping but not identical specificities of the domains. Specifically, peptides selected by the SH3C domain can host a proline (which can exclusively be substituted by a valine) in the position preceding the arginine of the core PxxxPR motif, whereas the other SH3 domains can tolerate polar and charged residues (Kurakin, Wu et al. 2003). This differential binding has also been observed not only with peptides but with natural partners as well, as it has been shown by pull-down or assays where different authors have noticed the lowest affinity to c-Cbl to be displayed by SH3C, as well as no appreciable binding between BLNK or CD2 and SH3C (although there was clear binding to SH3A and SH3B) (Kurakin, Wu et al. 2003; Tibaldi and Reinherz 2003). Another feature that distinguishes SH3C from the other CIN85 SH3 domains is that it is the least cross-reacting toward peptides selected by the other two domains, supporting the idea of overlapping but distinct specificities of CIN85 SH3 domains.

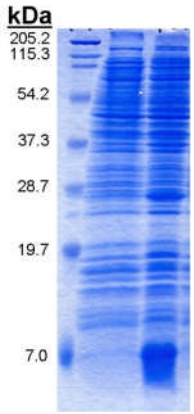


Figure V.9. Overexpression of CIN85-SH3C$_{270-328}$His$_6$. The CIN85-SH3C construct was expressed in *E. coli* BL21(DE3) strain at 25ºC, induced with 0.1mM IPTG. 17% SDS-PAGE of total cells: lane 1, molecular weight marker; lane 2, pre-induced and lane 3, induced. Gel was stained with Coomassie Brilliant Blue.

The SH3C$_{270-328}$ construct was successfully expressed in *E. coli*/BL21(DE3) with LB broth at 25ºC for 7 hours of 0.1mM IPTG induction (Figure V.9). It was purified following a similar strategy as followed for SH3B. No addition of protease inhibitors in the lysis buffer was used since the SH3 domains are predicted to be globular and their short sequence length is not expected to contain many exposed sites susceptible for bacterial proteolysis. However, as a precaution and to avoid activation of proteases, the

cell lysate was kept on ice or at 4ºC at all times prior to His-tag selection, which was carried out inmediately after cell lysis.

Briefly after sonication, the clear lysate was subjected to a His-tag selection in a nickel charged HiTrap chelating column, and bound protein was washed with a buffer containing 50mM imidazole so that waste proteins readily separated from the target protein (first eluted peak in the chromatogram, see Figure V.10). For protein elution a 50 to 500mM Imidazole gradient was performed. We recovered a large amount of a partially clean sample of SH3C, which eluted in a broad peak starting at 109mM Imidazole (Figure V.10).

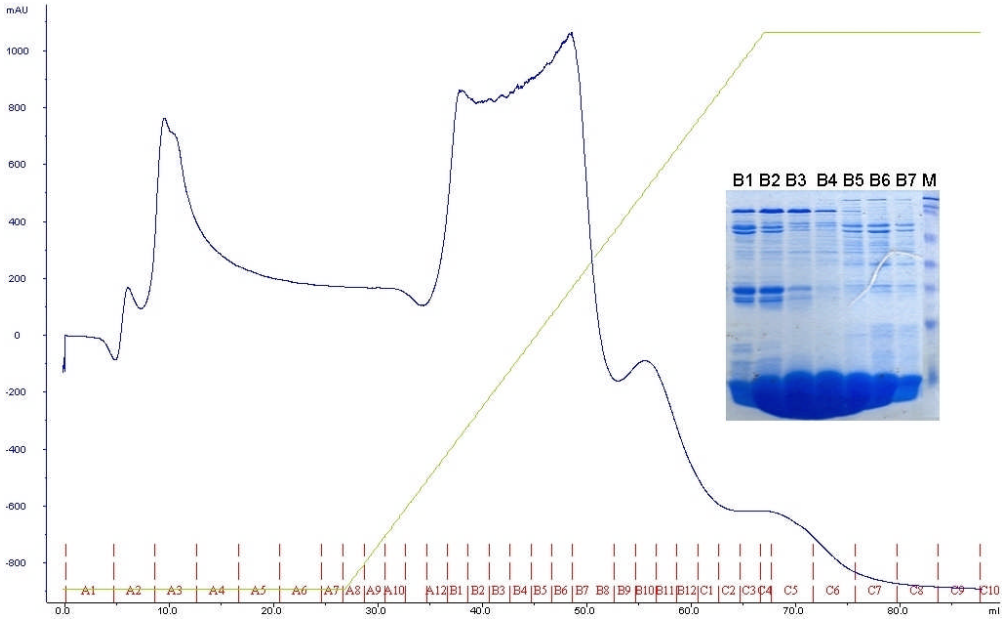


Figure V.10. Purification of -SH3C$_{270\text{-}328}$His$_6$ in a nickel charged Chelating HiTrap column. The blue spectrum indicates the UV_{280} absorption of eluted sample, and the green line indicates the elution buffer percentage (which, indirectly indicates the imidazole concentration from 50 to 500mM). The SH3C eluted in a broad peak between 109 and 266mM Imidazole concentration. The indented 15% polyacrylamide Coomassie Brilliant Blue stained SDS gel shows collected fractions of the elution peak where the SH3C protein was eluted together with some high molecular weight contaminants.

In order to further purify the sample, a second anion exchange chromatography step was performed. Given that imidazole contributes to increase the ionic strength of the solution, we diluted the collected sample to a calculated imidazole concentration of 20mM, previously to loading into the Q FF HiTrap column so that sample would bind to the resin. Most of the middle molecular weight contaminants did not bind to the column, and the SH3C protein eluted in a broad peak between 18 and 49mS/cm of conductivity (see Figure V.11).

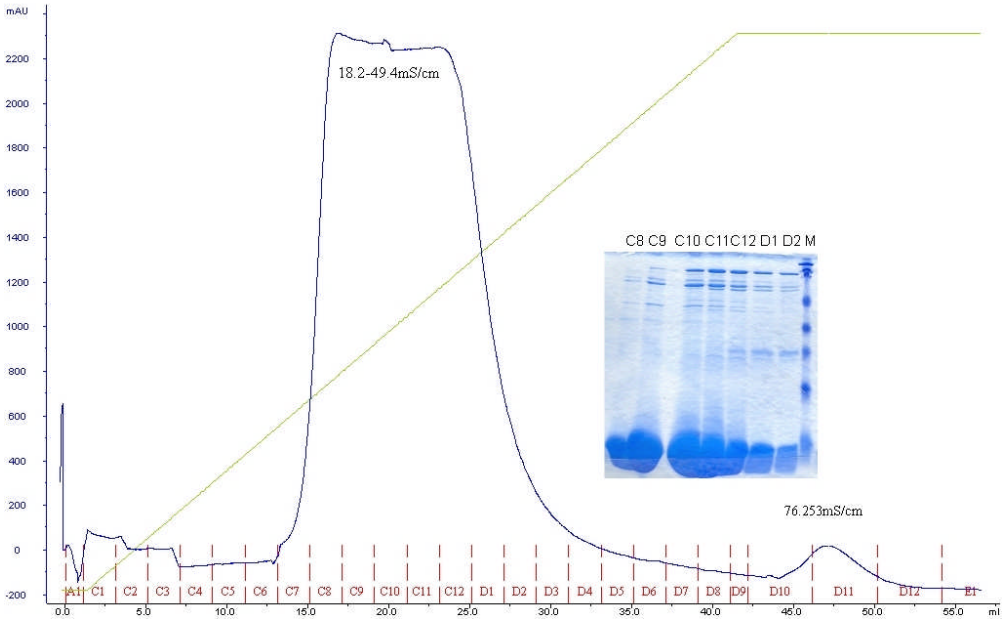


Figure V.11. Purification of CIN85-SH3C$_{270\text{-}328}$His$_6$ in a Q FF HiTrap column. The blue spectrum indicates UV$_{280}$ absorption of the eluted sample, and the green line shows the elution buffer percentage (which indirectly indicates the NaCl concentration, from 0 to 1M). CIN85-SH3C$_{270\text{-}328}$His$_6$ elutes in a broad peak starting at 18.2mS/cm. The indented 15% polyacrylamide SDS-PAGE stained with Coomassie Brilliant Blue, shows elution peak corresponding to SH3C elution.

The slightly purer sample was then subjected to a size exclusion chromatography in a Superdex75 column. A very pure protein sample that behaved as a monomer was collected. CIN85-SH3C displayed an apparent molecular weight of 8.6kD (Figure V.12), which is in agreement with the sequence calculated weigth (8.1kD).

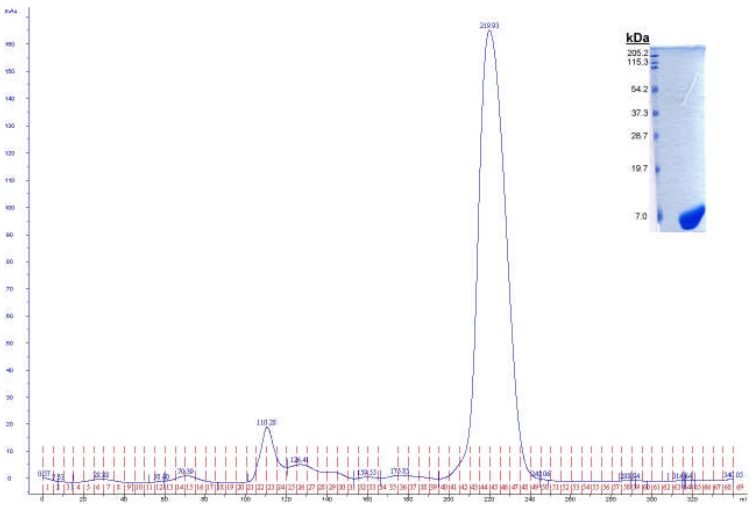


Figure V.12. Purification of CIN85-SH3C$_{270\text{-}328}$His$_6$ by size exclusion. Elution of CIN85-SH3C$_{270\text{-}}$328His$_6$ in a Superdex 75 26/60 gel filtration column at a volume of 219.93ml, which corresponds to a molecular weigh of 8.6kD that agrees with the theoretical weight of the monomeric SH3C (8.1kD). Analysis of the elution peak in a 15% SDS-PAGE shows high purity of the eluted sample.

Mass spectrometry analysis showed a spectrum similar to the one obtained for SH3B. The molecular weight measured differed from the calculated one in 140Da. The initiation methionine (formyl-methionine) has a theoretical weight of 161Da. This weight difference could be accounted for the loss of the formyl-methionine, as it has been previously described in the literature (Sherman, Stewart et al. 1985; Ben-Bassat, Bauer et al. 1987), together with the addition of a sodium atom (23Da) present in the buffer.

Circular dichroism studies in SH3C displayed a spectrum similar to the one observed for the SH3B, correlating with published CD spectra of correctly folded SH3 domains (Figure V.6B and C, and Figure V.13A). Apart from measuring the ellipticity changes at the far UV spectrum, a thermal denaturation curve at a 222nm wavelength was performed. CD is extraordinarily sensitive to changes in conformation. Hence, conformational changes upon ligand binding, or protein denaturation, for instance, can be followed. The denaturation of proteins is always accompanied by CD changes that indicate the loss of secondary structure and the enhancement of the random-coil spectral components. By plotting ellipticity at a particular wavelength as a function of the denaturing conditions, in this case temperature, one can follow denaturation of the protein.

Thermal denaturation of CIN85-SH3C$_{270-328}$His$_6$ yielded a sigmoidal transition reflecting loss of secondary structure when monitored by far-UV CD, as seen in Figure V.13B. A wavelength of 222nm was chosen to monitor the thermal melts because maximum differences in CD signal were seen at this wavelength. A cooperative process of denaturation occurred between 35 and 60ºC, with a melting temperature of approximately 45ºC (Figure V.13B).

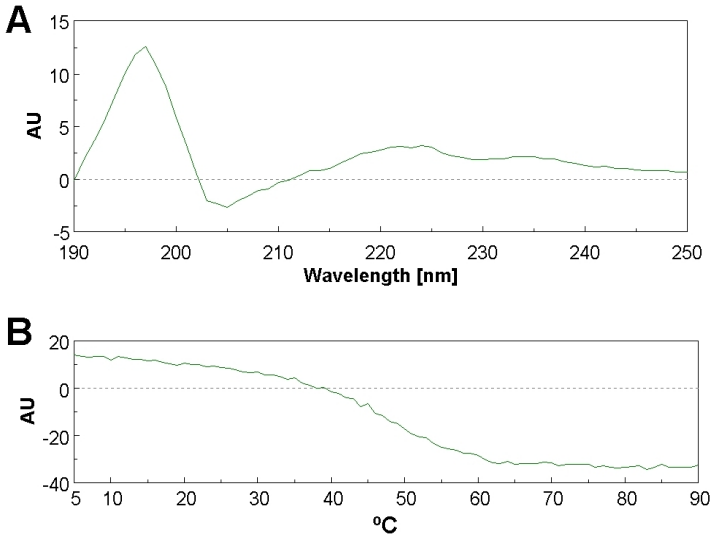


Figure V.13. Circular Dichroism (CD) spectra of CIN85-SH3C$_{270-328}$His$_6$. (A) The spectrum corresponds to ellipticity changes along the far UV spectrum. **(B)** Thermal stability curve of CIN85-SH3C$_{270-328}$His$_6$, with normal two folding behavior monitored by ellipticity at 222nm, where we observe the melting temperature at around 45ºC.

Previous knowledge of SH3 domain structure, allows us to search for specific secondary structure elements in the NMR spectrum of CIN85-SH3C$_{270-328}$His$_6$. As mentioned previously, SH3 domains fold into a β-barrel build of two antiparallel β-sheets. The presence of such antiparallel β-sheets was inferred in the

NMR analysis, with chemical shifts between 6.5 and 5.0 ppm characteristic of such structural elements (Figure V.14.) (Viguera, Arrondo et al. 1994; Viguera, Martinez et al. 1994).

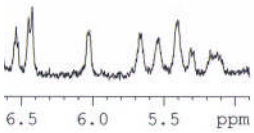


Figure V.14. Detail of the Monodimesional Nuclear Magnetic Resonance (NMR) spectra of the CIN85-SH3C$_{270-328}$His$_6$ protein construct between 6 and 5.5 ppm. Peaks in this region indicate the presence of β-sheet folded structure.

Provided that results from Mass spectrometry, CD and NMR analysis indicated evidence for the correct folding state of the SH3C construct in a homogesneous sample, the pure protein was used for crystallization trials.

1.1.2.1. CIN85-SH3C binds to c-Cbl peptides

In order to evaluate the biological activity of the newly designed SH3C construct and c-Cbl peptide, we wanted to verify their binding by fluorescence spectrometry and compare it with the published data.

To our convenience, there are only two tryptophans in the CIN85-SH3C sequence, which are located consecutively, and one of them (Trp307) is predicted to be directly involved in SH3:polyproline binding. Changes in Trp307 environment upon interaction with the ligand are thus expected to be reflected in fluorescence emission.

Fluorescence was hence used as a spectroscopic probe to monitor the change in environment of tryptophan 307 from the binding core domain and its adjacent tryptophan, Trp306, upon peptide binding. The affinity of CIN85-SH3C$_{273-326}$His$_6$ for the synthetic peptide c-Cbl$_{810-834}$ reported by Kowanetz and co-workers was of 14μM (Kowanetz, Szymkiewicz et al. 2003). When binding affinities of our CIN85-SH3C construct to both a long (c-Cbl$_{810-834}$) and a short (c-Cbl$_{820-830}$) c-Cbl peptide were measured, we obtained measurements of a dissociation constant six times lower than the published data in the case of the long peptide (2.4 ± 0.3μM), and 23-fold lower with the newly designed peptide (0.6 ± 0.06μM) (Figure V.15).

Calculation of binding stoichiometry and affinity depends directly on sample concentration. Hence, accurate measurement of the protein and peptide sample concentration is critical for data interpretation. The polyproline c-Cbl derived peptide sequences do not include any aromatic amino acid, so that any precise calculation of the peptide concentration cannot be achieved by spectrometric absorvance measurement at 280nm wavelength. Instead, the lyophilized peptide was weighted in an analytical balance and appropriate buffer volume added to achieve the desired stock concentration. This method accumulates a series of experimental errors in each step, such as weighting errors and pipetting. Therefore, this is the most plausible explanation for the affinity differences observed in c-Cbl$_{810-834}$ and c-Cbl$_{820-830}$ binding to CIN85-SH3C.

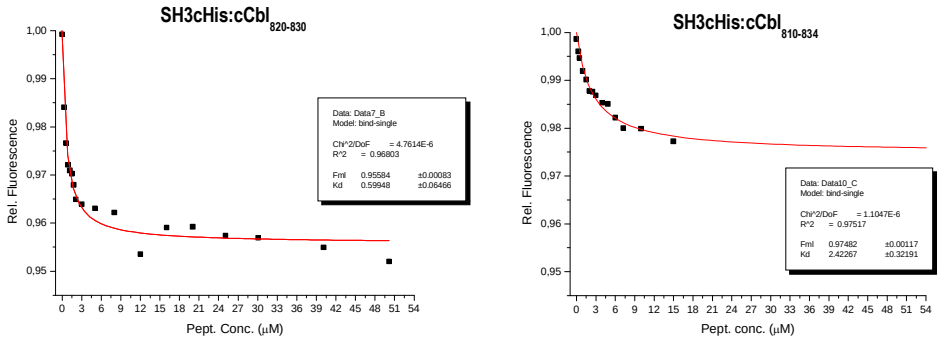


Figure V.15. Binding of c-Cbl peptides to CIN85-SH3C$_{273-326}$His$_6$ construct monitored by collisional quenching of CIN85-SH3C fluorescence. Change in the fluorescence intensity of CIN85-SH3C as a function of c-Cbl$_{820-830}$ (left) and c-Cbl$_{810-834}$ (right) concentration, corrected for the effect of dilution on addition of the buffer. The solid red line is the calculated curve using the best-fit parameters to equation $F_{rel}=(1+K[L]F_{ML})/(1+K[L])$. The best values for stoichiometry, binding constant, and enthalpy are shown in the box. A decrease in the relative fluorescence is observed upon peptide addition in both cases. To better illustrate the different affinities between these two titrations the x and y axes are shown on the same scale.

It is important to highlight the observation that most of the published fluorescence spectrometry studies of SH3 domain binding to polyproline sequences show an increment in the fluorescence when the peptide binds to the SH3, as shown by De Filippis and co-workers and Kaneko and co-workers (Figure V.16). Surprisingly, in our case there is a quenching effect upon peptide binding.

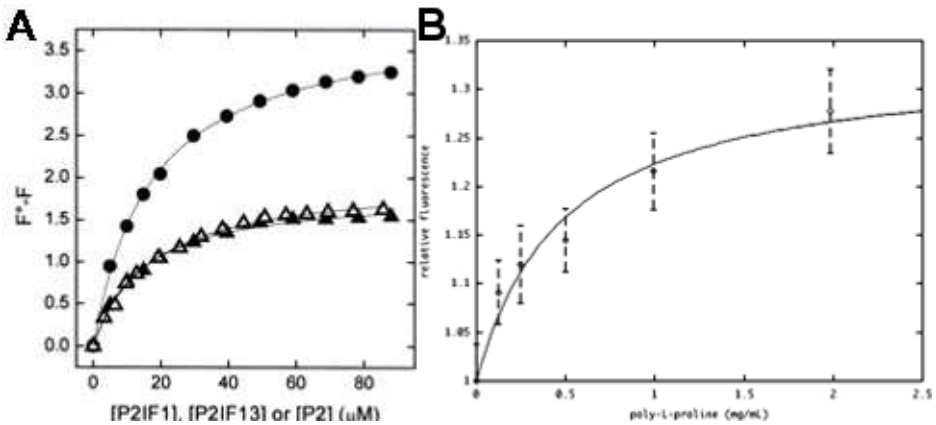


Figure V.16. Published fluorescence data of other SH3 domains binding to polyproline peptides. (A) Myo3-SH3 binding to IF analogs: P2IF1 (•-•), P2IF13 (▲-▲) and an unmodified P2 peptide (△-△). Continuous lines represent the best fit of the data points to $\Delta F=F^o-F=(\Delta F_{max}\cdot[L])/(K_D+[L])$ (De Filippis, Draghi et al. 2007) and **(B)** fluorescence change of STAM2 SH3 domain by poly-L-proline peptide binding (Kaneko, Kumasaka et al. 2003). The change in fluorescence upon peptide binding in both examples increases, as opposed to CIN85-SH3C binding to c-Cbl peptides.

Additionaly, we were able to co-purify a His/GST-tagged c-Cbl peptide (His/GST/c-Cbl$_{820-830}$, in pGATEV vector) with an untagged SH3C CIN85 construct (CIN85-SH3C$_{270-328\ taa\ tag}$, in pET-24d vector) by means of binding of the c-Cbl

peptide to a GSTrap column and then the SH3C to the peptide. Presence of both proteins in sample solution was confirmed by SDS-PAGE (Figure V.17) and mass spectrometry. From this result we could conclude that binding of both proteins was possible, although no affinity constant or stoichiometry can be estimated from such experiments. Crystallization experiments were set with this sample and crystals were obtained. Crystal diffraction was very poor and no data processing was possible.

Figure V.17. Purified CIN85-SH3C$_{270\text{-}328\ taa\ tga}$ (middle band) and c-Cbl$_{820\text{-}830}$ after tag cleavage (lower band). A small fraction of the Tev protease used for tag cleavage is present in this sample **(top band).**

1.1.2.2. Structure of CIN85-SH3C$_{270\text{-}328}$His$_6$

Un-complexed CIN85-SH3C$_{270\text{-}328}$His$_6$ was crystallized at 4ºC by the vapor diffusion method, in 1µl drops of a 1:1 mix of protein solution at 11.4mM (92.7mg/ml) and three different crystallization solutions: 0.1M Acetate pH 4.5, 20% PEG 3000; 0.1M Citrate pH 5.5, 2.0M $(NH_4)_2SO_4$ and 0.1M CHES pH 9.5, 30% PEG 3000 (Figure V.18).

Crystal formed in the second condition appeared to be the best candidate for its reproduction. Optimization of the crystal was carried out with different $(NH_4)_2SO_4$ concentrations and different citrate pH's (100mM final concentration). Crystals used for the X-ray diffraction were produced in 1.8M $(NH_4)_2SO_4$, 100mM Sodium Citrate pH 6.0 and stepped soaked prior to crystal diffraction into a cryo solution of reservoir solution containing 15 and 20% Glycerol. A total of 80 images of 1º rotation with 15 minute exposure each were collected with a maximum resolution of 2.0Å.

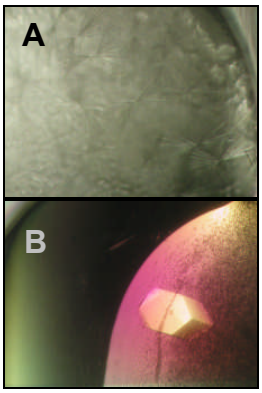


Figure V.18. CIN85-SH3C$_{270\text{-}328}$His$_6$ protein crystals. (A) Needle shaped crystals grown in 0.1M Acetate pH 4.5, 20%PEG 3000. **(B)** Single crystal of CIN85-SH3C$_{270\text{-}328}$His$_6$ grown in 0.1M Citrate pH 5.5, 2.0M $(NH_4)_2SO_4$.

Data reduction was undertaken with Mosflm software and Scala. The first 55 images were used to ensure good statistics. Standard deviation corrections were applied to reach an R_{sym} of 54% (see Table V.1). We reached a completeness of 95.4% at a resolution of 2.0Å (see Table V.1). The structure was subsequently solved in the $P3_121$ space group by molecular replacement (MR) from a molecule of the Sem5 C-terminal SH3 domain (chain A in PDB 1SEM) using the program Molrep (Collaborative Computational Project 1994). A cell volume of 138951Å^3 could only fit one molecule in the asymmetric unit (Matthew's coefficient 2.94, with a calculated solvent content of 59%).

A resolution range of 26.8-2.2Å was used all throughout the molecular replacement process. The integration radius for the rotation function was 20.16Å, corresponding to 60% of the longest dimension of the model (model dimensions: 33.8 x 31.0 x 26.5Å) and it is 1.9 times larger than its radius of gyration (10.4Å).

A good correlation factor of 43 and low R-factor of 58.2% with a contrast of 4.4 to the next molecular replacement solution indicated the correct solution. Rigid-body refinement of the top solution resulted in a correlation coefficient of 42 and an R value of 50.5%.

Automated model building was carried out with the program ARP/wARP (Zwart, Langer et al. 2004) included in the CCP4 program suite (Collaborative Computational Project 1994), with which 98% of the sequence was covered, (only the first alanine residue and the last four histidine residues from the vector sequence were missing). A high quality model was obtained with an R-factor (which measures the agreement between the crystallographic model and the experimental X-ray diffraction data) of 21% and 24% of R-free (which is used to validate the R-factor value) for a random 5% of all data.

Because the resolution of the data obtained by X-ray crystallography is not high enough to define the atomic positions of a macromolecule with enough precission, the structure models need to be optimized. Different programs have been developed to subject models from the experimental data to sophisticated refinement procedures that apply constrains based on prior knowledge about the chemical and conformational properties of the atoms and the molecule.

Table V.1. Crystallographic data collection and refinement analysis of CIN85-SH3CHis structure

CIN85-SH3CHis	
Space group	$P3_121$
Unit Cell Dimensions (Å)	a=b=47.83, c=70.23
	α=β=90º; γ=120º
Data range (Å)	26.8-2.0
Observations (unique)	19366 (5991)
Completeness (%) (last shell)	95.4 (90.6)
R_{sym} [a] (last shell)	0.054 (0.032)
Reflections F>0 (cross validation)	5715 (275)
Non-hydrogen atoms (solvent molecules)	606 (58)
R_{cryst} [b] (R_{free} [c])	0.197 (0.248)
R.m.s. bond length (Å)	0.023
R.m.s. bond angles (º)	1.674

[a] R_{sym} is the unweighted R value on I between symmetry mates.
[b] $R_{cryst} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$
[c] R_{cryst} is the crossvalidation R factor for 5% of reflections against which the model was not refined.

The model of CIN85-SH3C generated with ARP/wARP was subjected to cycles of manual and automatic refinement with the programs Coot V0.1.2-Pre-1 (Emsley and Cowtan 2004) and Refmac5 (Collaborative Computational Project

1994) (from the CCP4 program suite (Collaborative Computational Project 1994)), respectively. Inspection of electron density map in the program Coot allowed building of the initial alanine and two more histidine residues in the His-tag, as well as the addition of water molecules up to a total of 58. Repetitive cycles of positional and isotropic thermal factor restrained refinement resulted in an R-factor of 19.7% and an R-free of 25.1% for data at 2.0Å resolution (see Table V.1).

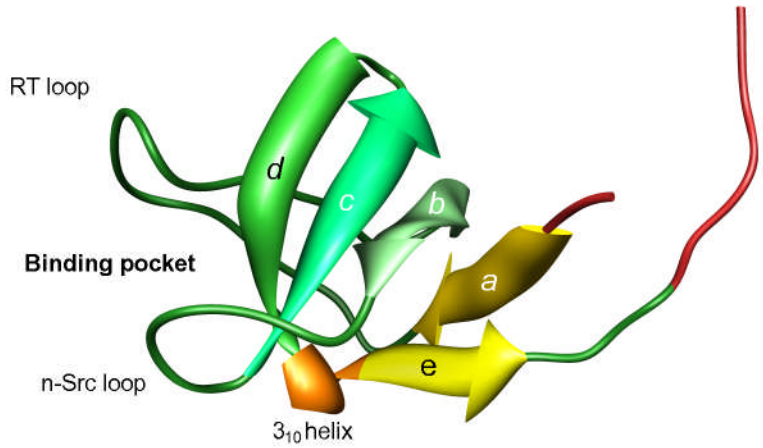


Figure V.19. Structure of CIN85-SH3C$_{270-328}$His$_6$. This cartoon representation shows a β-barrel structure characteristic of SH3 domains. Main structural elements, such as β-strands (a through e), 3_{10} helix, RT-loop, n-Src loop and binding pocket are indicated. Residues belonging to the vector sequence are shown in red.

The structure determination and refinement procedures might have biases that prone the structure models to systematic errors. Therefore, it is necessary to validate the quality of such models. There are a variety of tools available that measure the deviation of the geometric and conformational parameters of the model from those of other small molecules or high-quality structures of other macromolecules.

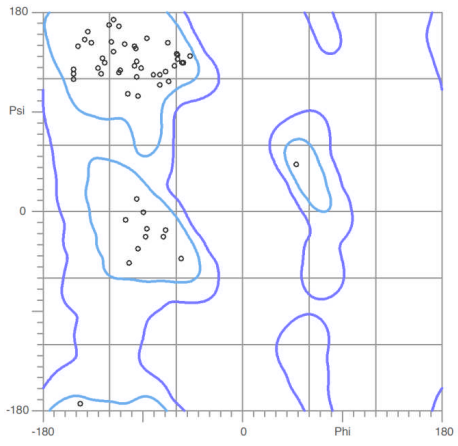


Figure V.20. Ramachandran plot of CIN85-SH3CHis$_6$ structure validation in MolProbity webtool. There are no residues in disallowed regions and 100% of them are in the favored regions.

We analyzed our CIN85-SH3C structure model with the web-based tool MolProbity (Lovell, Davis et al. 2003) (at http://molprobity.biochem.duke.edu). Evaluation with this tool showed all residues in the favored regions of the Ramachandran plot, which compares the backbone φ, ψ angle distribution with that of high-quality structures (See Figure V.20). No serius steric overlaps were detected, and correct geometry was confirmed.

The structure of CIN85-SH3C clearly displayed the characteristic β-barrel fold of SH3 domains formed by two antiparallel β-sheets perpendicular to each other (Figure V.19). This arrangement forms a hydrophobic pocket in the surface between the two β-sheets, where the polyproline peptide binds. The first β-sheet is formed by β-strands *a* (residues 270-274), *e* (residues 323-325) and first half of *b* (residues 293-295), while the second is comprised of β-strands *c* (a.a. 306-311), *d* (a.a. 314-319) and the second half of *b* (a.a. 296-298). It is worth mentioning that one of the most conserved positions in SH3 domains is a glycine (Gly316 in CIN85-SH3C) located in the middle of β-strand *d*. This fact is not common in β-strands, since glycines are usually unfavored. It has been proposed that the reason why this glycine is present at this position is because a bulkier residue would collide with side chains of residues from the RT-loop, although Larson and Davidson observe that position of this glycine adopts uncommon Φ/ψ[6] angles (Larson and Davidson 2000).

A torsion in β-strand *b*, known as a β-bulge, is what allows this strand to participate in both β-sheets. These β-bulges are secondary structures that occur mostly as irregularities in antiparallel sheets (only 5% of them have been found in parallel sheets) (http://65.54.170.250/cgi-bin/linkrd?_lang=EN&lah=10c153948089623929c773ea6ff7b706&lat=1086808043&hm__action=http%3a%2f%2fwww%2end%2eedu%2f%7etlin%2fprotein%2ehtm). The bulge in this sheet is due to hydrogen bond formation between one amino acid from one sheet and one from the other. This β-bulge is conserved in SH3 domains at a position corresponding to isoleucine 297 in CIN85-SH3C. Conserved hydrogen bonds are present between amido groups of this residue and the following and the carbonile oxygen of residue corresponding to $Glu308_{CIN85}$. Another bond between the amido hydrogen of the latter and the carboxylic oxygen from residue corresponding to $Asn298_{CIN85}$ is also present (Larson and Davidson 2000) (Figure V.21).

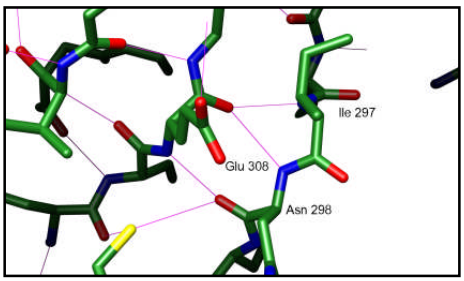


Figure V.21. Detail of the β-bulge in the CIN85-SH3C$_{270-328}$His$_6$ structure. Hydrogen bonds are represented by pink lines. Carboxylic oxygen from Glu308 main chain hydrogen bonds amide groups from Iso297 and Asn298 main chains and amide group of Glu308 with carboxylic oxygen of Asn289.

[6] Φ refers to the angle around C_α-N and ψ the angle aroundC_α-C_o. Positive values correspond to a clockwise turn of the angle (viewed from the C_α). (Φ,ψ)=(0,180), two carbonile oxygens are too close; (Φ,ψ)=(180,0), two amide groups are superposed; (Φ,ψ)=(0,0), one carbonile oxygen is superposed with an amide group. Glycines allow a broader angle range than other residues.

Several hydrogen bonds in the RT-loop (located between β-strands *a* and *b*) give a distinctive irregular antiparallel structure with β-sheet character observed in most SH3 domains (Gmeiner and Horita 2001). A number of recurrent interactions are observed in all SH3 domains analyzed by Larson and Davidson (Larson and Davidson 2000) that stabilize this loop, such as hydrogen bonds between $Glu285_{CIN85}$ and $Tyr278_{CIN85}$, and carbonyl oxygen from $Thr287_{CIN85}$ and the amido hydrogen from $Ala280_{CIN85}$. The hydrogen bond commonly found between the residue located five amino acids prior to $Glu285_{CIN85}$ (usually a lysine) and the residue 13 positions prior to it (occupied 60% of the time by an asparagine) (Larson and Davidson 2000) is not present in CIN85-SH3C, given that instead of the conserved asparagine, we find a proline (Pro277), and a glutamic acid (Glu290) in position of the conserved lysine. Proline is not a hydrogen donor, so that this hydrogen bond cannot be formed.

The point in which the first β-strand changes direction is where the RT-loop starts. Here, a hydrogen bond formed by a highly conserved residue is widely present in SH3 domains (Larson and Davidson 2000). In CIN85-SH3C, this hydrogen bond is estabilised between conserved residues Iso275 and Phe322 (see Figure V.22). It has been postulated that even though the residue corresponding to isoleucine in CIN85-SH3C is not directly implicated with the recognition of the polyproline motif, it can play a functional role in the context of full length protein complexes (Larson and Davidson 2000). In the case of CIN85, this position is occupied by a glutamic acid in SH3A and an alanine in SH3B.

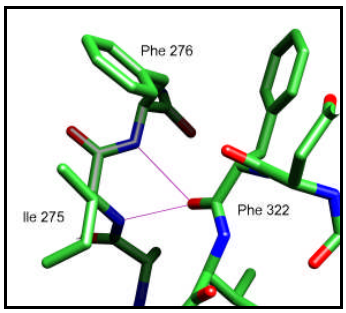


Figure V.22 Detail of the RT loop in the CIN85-SH3C$_{270\text{-}328}$His$_6$ structure. Hydrogen bonds are represented with pink lines. Conserved hydrogen bond between Ile275 and Phe322 is shown. Phenylalanines 276 and 322 that are predicted to interact directly with the polyproline peptide are labeled.

Although SH3 sequences are variable, all of them share a highly conserved common fold. In order for this to happen, certain residues need to be preserved. Many of them form a hydrophobic nucleus critical for the protein stability, which is comprised of a series of hydrophobic side chains particularly inaccessible to the solvent. Almost ten positions that are conserved in the SH3 domains have hydrophobic character and are packed in the center of the structure. These correspond to the following in the SH3C of CIN85: Val274, Leu286, Iso288, Val294, Leu296, Trp307, Phe318 and Val323, and also Lys273. Glycine 309 is considered as part of the nucleus, even though it does not have a hydrophobic character, given that it is always completely buried and contacts are established with side chains of the rest of the residues in the nucleus.

Other conserved hydrophobic residues are present in these structures, generating two hydrophobic pockets that, together with an acidic pocket located

between the RT and n-Src loops, are critical for the protein recognition. Aromatic residues Trp306 and Tyr278, and Pro319 in SH3C are the positions most conserved in SH3 domains and play an essential role in the binding to the peptide, given that they are situated in the center of the interaction. Residues corresponding to phenyalanines 276 and 322, glycine 305 and asparagine 321 interact directly with the peptide and thus are also determinant for the binding.

Analysis of the arrangement of the symmetric molecules in respect to the asymmetric unit shows that the histidine tag packs directly against the acidic pocket of the crystallographic symmetry related molecule, where the peptide was expected to bind (Figure V.23). There is a hydrogen bond between the amide groups of the first histidine and the conserved tryptophan 306, and between the hydroxyl group from the aspartic acid 303 side chain and the carboxylic oxygen of the same histidine. The second histidine forms four hydrogen bonds with Tyr278, Glu285 and Gln281 from the symmetric molecule (Figure V.23). All these residues are predicted to be essential for the binding of the ligand, so that the histidine tag could be competing with the peptide for the binding site and thus interfering in the crystallization of the complex.

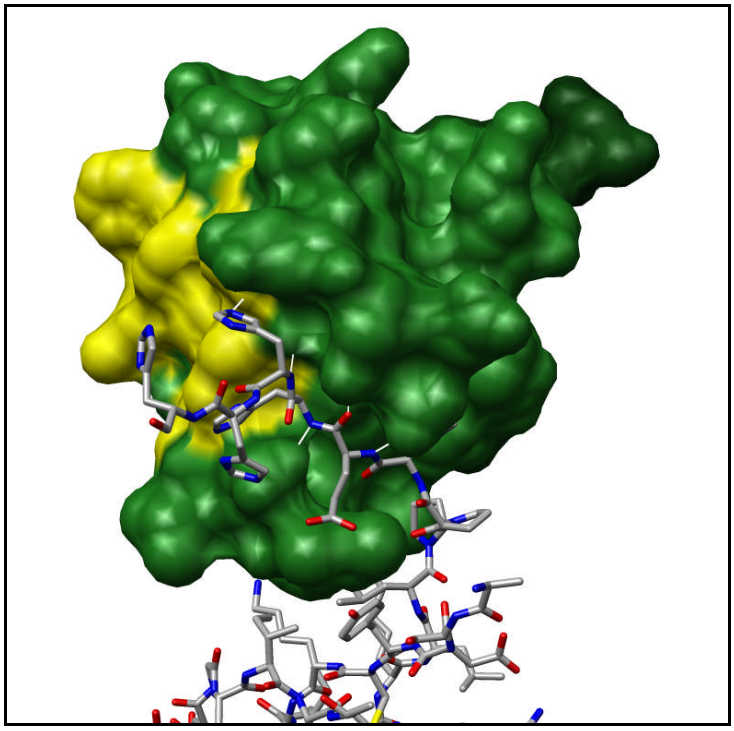
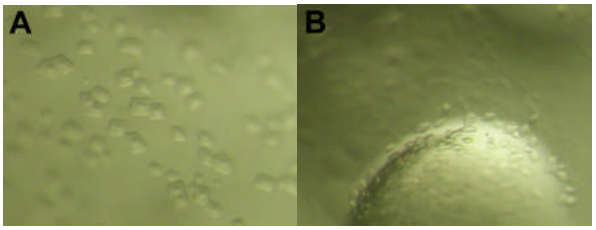


Figure V.23. Surface representation of the CIN85-SH3C$_{270-328}$His$_6$ structure. A symmetry related molecule surface is shown in sticks. Residues from the binding core, predicted to be involved in the peptide binding, are colored in yellow. The fraction of the histidine tag that could be modeled, shown in from the symmetrically related molecule, is pointing towards the binding surface.

We also set crystallization experiments of the different SH3C CIN85 constructs in presence of Cbl peptides of different lengths. We were able to obtain crystals at 4ºC that took over a year to appear, and were neither reproduceable nor big enough for X-ray diffraction (Figure V.24).

Figure V.24. Crystals obtained with a protein mix of CIN85-SH3C$_{270-328}$His$_6$ and Cbl-b$_{902-912}$ at 4ºC. Crystals were grown in **(A)** 30% PEG 4000, 100mM Na Citrate pH5.6, 200mM

Amoniumm Acetate and **(B)** 25% PEG 4000, 8% Isopropanol, 100mM Acetate.



1.1.3. Purification, characterization, crystallization and structure of CIN85-SH3A$_{2-58}$

In parallel to the SH3C structural studies, we tried the crystallization of the CIN85-SH3A.

High overexpression of the soluble CIN85-SH3A protein was achieved in Rosetta(DE3)pLys *E. coli* at 37ºC with 2xTY or SOC broth, induced with 1mM IPTG for 4hr (see Figure V.25).

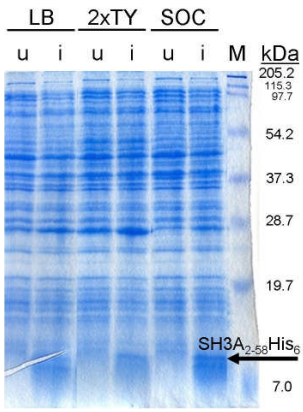


Figure V.25. Overexpression of CIN85-SH3A$_{2-58}$His$_6$ in *E. coli* Rosetta(DE3)pLysS strain. Expression was achieved at 37ºC in LB, 2xTY and SOC broths and induced with 1mM IPTG. 17% SDS-PAGE gel of total cells; lane 7, molecular weight marker (M); lanes 1 and 2: expression in LB, uniduced (u) and induced (i); lanes 3 and 4: expression in 2xTY, uniduced (u) and induced (i); lanes 5 and 6: expression in SOC, uniduced (u) and induced (i). Gel was stained with Coomassie Brilliant Blue.

The purification protocol used was modified slightly from the previous ones. Low concentration of imidazole (15mM) was also used in the sonication buffer to avoid binding competition of contaminants when loading the sample in the nickel charged chelating column.

In this case, two washing steps were performed: a first wash with 15mM imidazole and 400mM NaCl, which were reduced to 2mM Imidazole and no sodium chloride in the second washing step. The protein was subsequently eluted at 261mM Imidazole in a 2 to 500mM imidazole gradient (Figure V.26). The SH3A protein came out very concentrated but contaminated with high molecular weight proteins, so that an anion exchage step was required. As mentioned before, the sample needed to be diluted to a calculated Imidazole concentration of 20mM, prior to loading into the Q sepharose column. The protein was eluted at 17.5mS/cm conductivity in the anion exchange chromatrography step, which indicated similar global electrostatic properties as observed for CIN85-SH3C (Figure V.27). After a final purification step in gel filtration separation, maximum purity was achieved, and the protein was recovered at an elution volume of 70.72ml (Figure V.28), corresponding to a monomeric state (7.0kD) according to the column calibration (theoretical weight calculated to be 7.8kD).

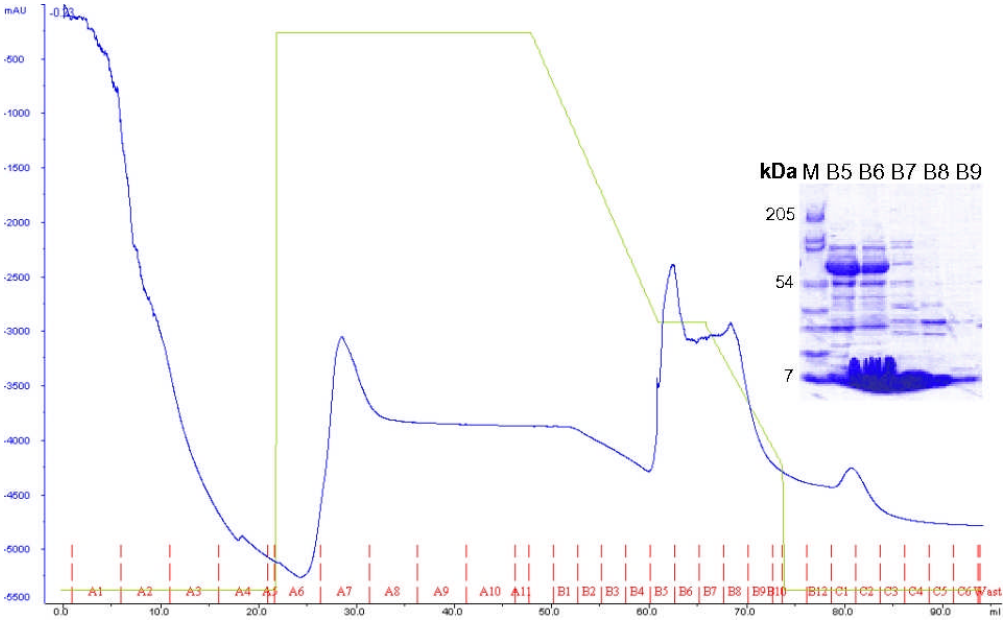


Figure V.26. Purification of CIN85-SH3A$_{2\text{-}58}$His$_6$ in a nickel charged Chelating HiTrap column. The blue line indicates UV$_{280}$ absorption of eluted sample, and the green line represents the percentage of buffer in pump B, which indirectly indicates the EDTA concentration (0-100mM). The SH3A is eluted in a broad peak with 261mM average Imidazole concentration. The indented polyacrylamide Coomassie Brilliant Blue stained SDS-PAGE gel, shows collected fractions of the elution peak where the SH3A protein was eluted together with some high molecular weight contaminants.

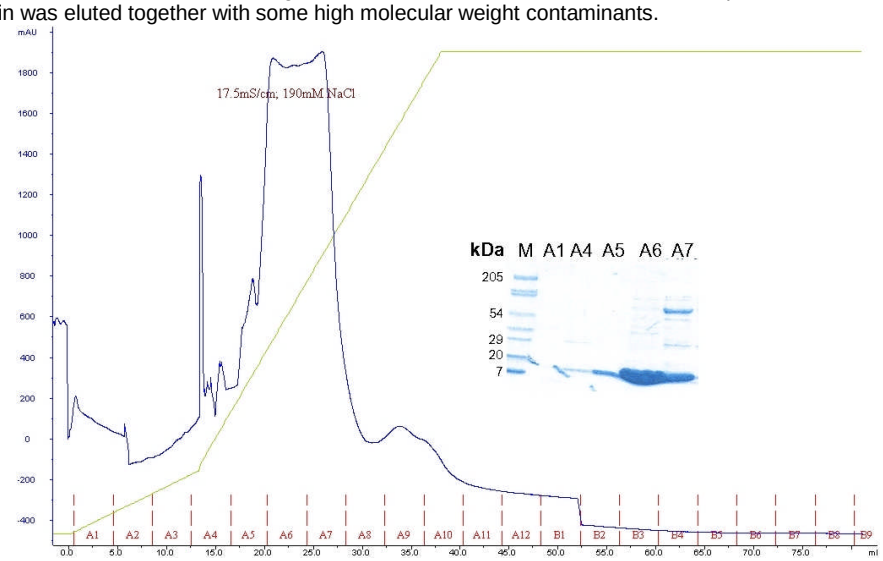


Figure V.27. Purification of CIN85-SH3A$_{2\text{-}58}$His$_6$ in a Q HP HiTrap column. The blue line indicates UV$_{280}$ absorption of eluted sample, and the green line shows the elution buffer percentage, which indirectly indicates the NaCl concentration (0 to 0.5M). CIN85-SH3A$_{2\text{-}58}$His$_6$ elutes in a broad peak starting at 17.5mS/cm, which indicates that this SH3A construct has similar superficial electrostatic characteristics as CIN85-SH3C$_{270\text{-}328}$His$_6$. The indented SDS-PAGE gel stained with Coomassie Brilliant Blue, shows elution peak corresponding to SH3A elution.

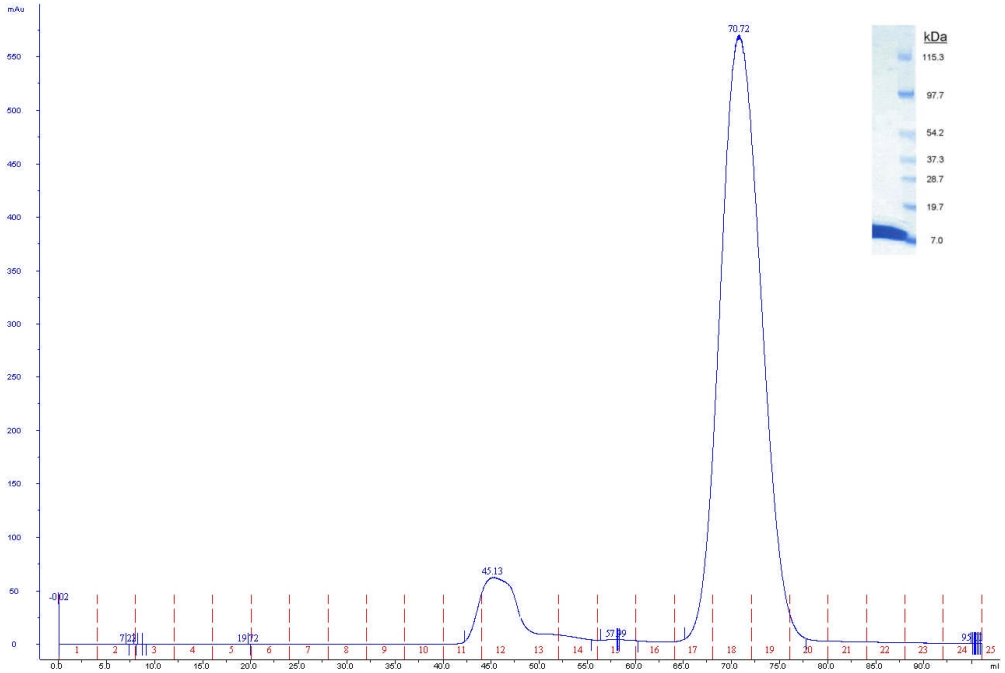


Figure V.28. Pure CIN85-SH3A$_{2\text{-}58}$His$_6$. Elution of CIN85-SH3A$_{2\text{-}58}$His$_6$ in a Superdex 30 16/60 gel filtration column at a volume of 70.72ml, which corresponds to a molecular weigh of 7.0kD and approximates to the theoretical weight of the monomeric SH3A (7.8kD). Analysis of the elution peak in an SDS-PAGE shows high purity of the eluted sample.

Spectrometry analyses were carried out for the pure sample as described previously for SH3B and SH3C. Mass spectrometry assay revealed a molecular weight 135Da lower than the theoretical weight of the CIN85-SH3A$_{2\text{-}58}$His$_6$ construct, which, as discussed for CIN85-SH3C$_{270\text{-}328}$His$_6$, could be accounted for the loss of the initiation methionine.

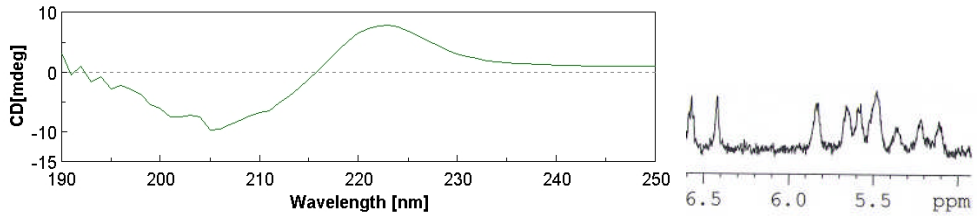


Figure V.29. Circular Dichroism (CD) spectra of CIN85-SH3A$_{2\text{-}58}$His$_6$ (A) and detail of the Monodimesional Nuclear Magnetic Resonance (NMR) spectra of the CIN85-SH3A$_{2\text{-}58}$His$_6$ protein construct between 6 and 5.5 ppm. Peaks in this region indicate the presence of β-sheet folding structure **(B)**.

CD and NMR 1-D spectra showed similar patterns as shown previously for SH3B and SHC (Figure V.6, Figure V.8, Figure V.13, Figure V.14, and Figure V.29), so that homogeneity and folding state of the sample was also confirmed. This protein sample seemed to show appropriate conditions to be used for crystallization assays,

since the homogeinity and purity of the sample is crucial for the acquisition of well-ordered crystals.

We also aimed to characterize the binding of the SH3A domain to the Cbl-b peptide by means of energetic terms. We wanted to obtain data on the stoichiometry of the binding, for which purpose we used titration microcalorimetry method to do the binding measurements.

To achieve accurate calculated affinity and stoichiometry values, we used the YCbl-$b_{903-912}$ peptide, which was tagged with a tyrosine residue in the N-terminus, so that concentration could be empirically determined by means of UV_{280} absorvance measurement (see Materials and Methods section 4.2).

CIN85-SH3$A_{2-58}$$His_6$ domain showed a clear binding to the Cbl-b peptide, with an affinity of 11.06μM (K = 9.04E4 ± 9.4E3M) (Figure V.30), which differs by more than four times with the results obtained in the fluorescence studies of SH3C:c-Cbl binding. This may indicate a higher affinity of SH3C to c-Cbl than that of SH3A to Cbl-b. In spite of that, the affinity values are in agreement with results published by Kowanetz and co-workers (Kowanetz, Szymkiewicz et al. 2003) so that a more plausible explanation to such difference could be that values are more reliable in this assay, given that the peptide concentration could be measured with higher precission.

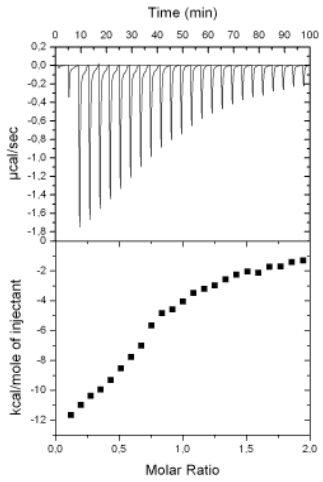


Figure V.30. Isothermal titration calorimetry measurements of complex formation between CIN85-SH3A (50μM) and the Cbl-b peptide (545μM). The upper panel shows the raw data of each titration. The optimal fits for these two titrations give a K_D of 11.6μM and an N of 0.75 in a one-site binding model.

We repeatedly obtained a non-equimolar stoquiometry, of 1:0.75 protein to peptide when data were analyzed in a one site system (Katrin Rittinger's lab reported a 1:0.61 soichiometry in the same experiment). This result opened a new door for the recognition of polyproline peptides by SH3 domains.

The fact that the proteins were able to bind the Cbl peptides also increased our hope on the acquisition of crystals of the SH3:Cbl peptide complexes.

Subsequently, the pure CIN85-SH3A protein sample was concentrated by ultracentrifugation with Amicon Ultra or Vivaspin filter devices with a 3000 or 5000Da membrane pore. Crystallization assays of the uncomplexed protein were set at 10.3mM (81mg/ml) protein concentration in 96-well plate format under the sitting drop vapor diffusion method. All crystallization commercial screens available in the laboratory were set at 22ºC and 4ºC with 1.6μl drops and initial crystal hits were observed at D4 and B9 Hampton conditions (0.1M Citrate pH 5.6, 20% PEG 4000, 20% Isopropanol and 0.2M Magnesium Acetate tetrahydrate, 0.1M Sodium

cacodylate pH 6.5, 30% 2-Methyl-2,4-Pentanediol) and at C9 Wizard I condition (0.2M Lithium sulfate, 0.1M CAPS pH 10.5, 2M Ammonium sulfate).

No crystal optimization was needed, and a single crystal from original Hampton condition B9 was used for solution of the structure. The crystal was washed in stabilization solution with 20% Ethylene Glycol (20% Ethylene Glycol, 80mM Citrate pH5.6, 16% PEG 4000, 16% Isopropanol) and immediately flash cooled under a nitrogen gas stream source. A dataset of 331 images of 0.5º rotation each, with 2 minutes of exposure was collected.

Probably due to some crystal cracking during manipulation, two crystal lattices could be distinguished in the diffraction patterns, and a high mosaicity (1.5º) was observed when the data was processed (Figure V.31). The crystal showed symmetry of a primitive tetragonal crystal, $P4_32_12$ and maximum resolution of 2.2Å. CIN85-SH3C coordinates were used as a search model for obtaintion of the theoretical phases. With a cell volume of 166814Å^2, only one molecule in the asymmetric unit could be fit (Mathew's coefficient of 2.81 and a calculated solvent content of 56%). The correlation coefficient for a single solution from AmoRe after rigid body refinement from the translation search was 35.6 and the R value was 51.4%.

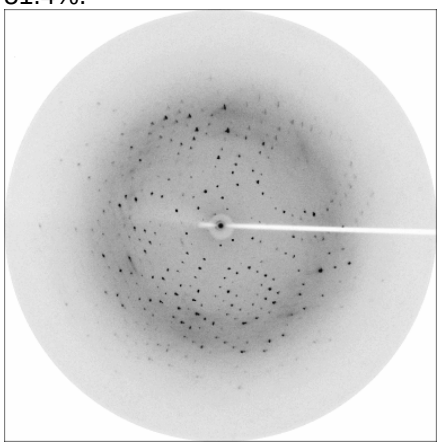


Figure V.31. Image of CIN85-SH3A$_{2\text{-}58}$His$_6$ crystal diffraction. This image corresponds to a diffraction exposure of 2 minutes in a 0.5º oscillation in phi, with maximum resolution at 2.2Å. A very mosaic diffraction pattern was detected as well as two crystal lattices diffraction patterns from the crystal cracking.

The structure showed the same β-barrel folding described for the SH3C, where the hydrophobic nucleus has isoleucines in place of Lys273, and Val294 in SH3C (Iso5 and Iso26 in SH3A), and an asparagine instead of Leu296 (Asn28). All residues from the binding core are conserved in SH3A (Figure V.32).

```
CIN85-SH3A  MVEAIVEFDYQAQHDDELTISVGEIITNIRKE--DGGWWEGQINGRRGLFPDNFVREIKK
CIN85-SH3C  -DYCKVIFPYEAQNDDELTIKEGDIVTLINKDCIDVGWWEGELNGRRGVFPDNFVKLLPP
            . * * *:**:******. *:*:* *.*:  * *****::*****:******: :
```

Figure V.32. Alignment of CIN85 SH3A and SH3C sequences. Conserved residues from the hydrophobic nucleous are indicated with pink boxes and conserved residues in the binding core are indicated with green boxes.

Nevertheles, in this crystal, electron density for the His-tag could only be detected for the first histidine residue, probably owing to its flexible nature. The beginning of the tag did not seem to be pointing towards the direction of any of the binding pockets of the crystallographically symmetry related SH3 domains though.

1.1.3.1. Structure of CIN85-SH3A$_{2-58}$:Cbl-b$_{902-912}$ complex

c-Cbl had been shown to bind preferentially to the N-terminal SH3 domain of CIN85 (Kowanetz, Szymkiewicz et al. 2003; Tibaldi and Reinherz 2003). Thus, the structure of CIN85-SH3A in complex with a Cbl peptide did seem as a good objective to pursuit.

Observations obtained from the crystallization of the constructs of the CIN85 SH3C and SH3A domains, and data obtained in the laboratory in related projects, allowed us to infer that the decrease in the number of residues foreign to the original protein sequence, together with the increase in the protein concentration favors its crystallization.

In order to avoid the possible interference of the histidine tag and the process of removal of any affinity tag used for the purification process, we decided to express the protein untagged and purify it by the traditional ammonium sulfate step precipitation method. For that purpose, the CIN85-SH3A was cloned in the pET-29b vector with a stop codon at the end so that the Open Reading Frame (ORF) did not include the His-tag.

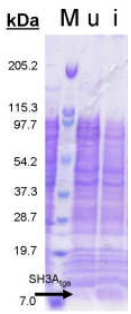


Figure V.33. Overexpression of CIN85-SH3A$_{2-58\ tga}$. The CIN85-SH3A$_{2-58\ tga}$ was grown in *E. coli* Rosetta(DE3)pLysS strain at 37ºC in 2xTY broth and induced with 1mM IPTG. A 4-22% SDS-PAGE of total cells is shown; lane 1: molecular weight marker (M); lane 2, uninduced (u) and lane 3: induced (i). Gel was stained with Coomassie Brilliant Blue.

Optimal expression of the pET29b/CIN85-SH3A$_{2-58\ tga}$ clone was achieved in 4 to 6 hours in Rosetta(DE3)pLysS *E. coli* cells induced with 1mM IPTG, in 1L cultures with 2xTY medium (see Figure V.33). Cell pellets were resuspended in Tris buffer (see Table IV.5) with 5ml of buffer per gram of cells, and frozen under liquid nitrogen. These cell pellets were thawed under ice and sonicated for five minutes and membrane traces were discarded by high speed centrifugation (16100 xg). Due to the extremely high solubility of the protein, we were able to salt-out the majority of the protein contaminants at 70% w/v ammonium sulfate concentration (see Figure V.34). The soluble fraction was further purified and partially desalted in a hydrophobic interaction chromatography step with a 50 to 0% $(NH_4)_2SO_4$ gradient (Figure V.35). A very pure fraction was collected and subsequently concentrated with an Amicon 5kD filter to a volume between 1.5 and 2ml for loading into a Superdex 75 16/60 column. The protein behaved as a monomer in the gel filtration chromatography (Figure V.36).

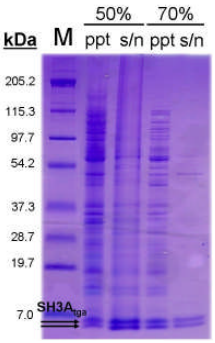


Figure V.34. 4-23% polyacrylamide gradient SDS-PAGE shows the $(NH_4)_2SO_4$ precipitation of CIN85-SH3A$_{2-58\ tga}$. Lane 1: Molecular weight marker (M); Lane 2: sample precipitated (ppt) with 50% ammonium sulfate; Lane 3: supernantant (s/n) from 50% ammonium precipitation; Lane 4: sample precipitated (ppt) with 70% ammonium sulfate; Lane 5: supernantant (s/n) from 70% ammonium precipitation. Sample recovered after salting out with 70% ammonium sulfate is visibly pure.

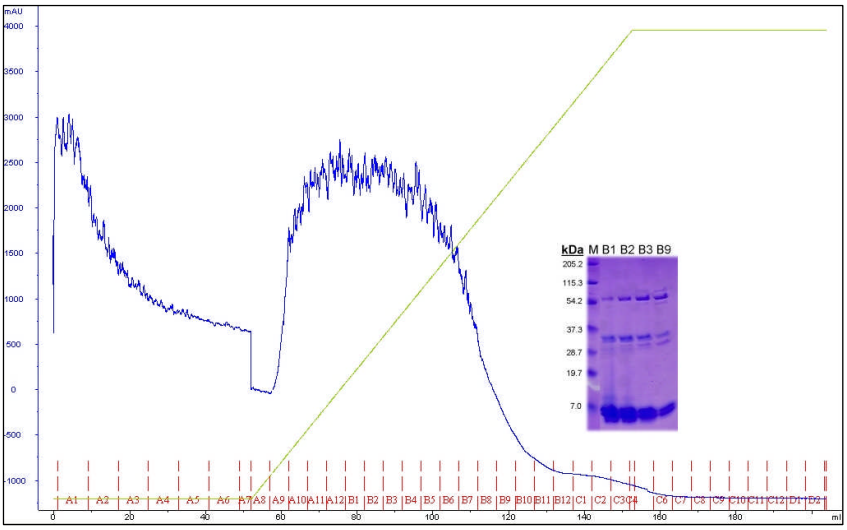


Figure V.35. Purification of CIN85-SH3A$_{2\text{-}58\ tga}$ by hydrophobic interaction chromatography. The blue line in the chromatogram indicates UV_{280} absorption of eluted sample, and the green line represents the percentage of buffer, from which the ammonium sulfate concentration can be inferred. The SH3A is eluted in a broad peak starting at 40% $(NH_4)_2SO_4$ concentration. The indented 4-23% polyacrylamide Coomassie Brilliant Blue stained SDS-PAGE gel, shows main collected fractions of the elution peak where the SH3A protein was eluted together with some high molecular weight contaminants.

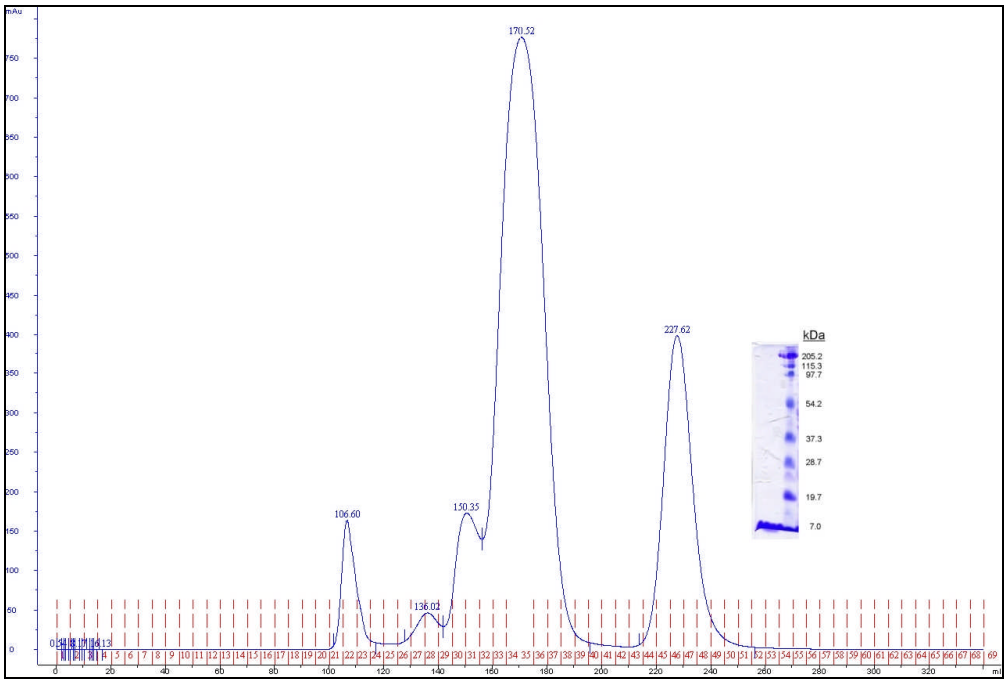


Figure V.36. Pure CIN85-SH3A$_{2\text{-}58\ tga}$. Elution of CIN85-SH3A$_{2\text{-}58\ tga}$ in a Superdex 75 26/60 gel filtration column at a volume of 227.62ml, which corresponds to a molecular weigh of 10.6kD and approximates to the theoretical weight of the monomeric SH3A (6.8kD). Analysis of the elution peak in 4-23% polyacrylamide gradient SDS-PAGE shows high purity of the eluted sample.

It is worth mentioning that a large amount of the expressed SH3 protein precipitated with the *E. coli* contaminants in the bulk salt precipitation step, and therefore we obtained a much lower yield compared to the affinity chromatography purifications, with a final harvested amount of aproximately 2mg of protein per culture Litre. However, the high purity of the sample was optimal for macromolecular crystallization.

Pure protein was concentrated to more than 7.4mM (50mg/ml) and mixed with the different proline rich peptides in ratios that varied from 1:2 to 1:3. Commercial crystallization screens were set at 22ºC for crystallization trials.

Crystals of CIN85-SH3A$_{2\text{-}58\ tga}$:Cbl-b$_{902\text{-}912}$ complex were obtained with a mixture of the pure protein and the peptide in a 1:2.5 protein:peptide ratio. Final concentration of the protein in the 1µl drops (0.5:0.5µl protein:reservoir solutions) was calculated to be of 7mM (47mg/ml).

In the initial crystallization Wizard screening by the sitting drop vapor diffusion method, 18 conditions were found to give crystals (Figure V.37). The large number of crystallization conditions allowed us to select for the best alternative, so that reproduction of crystals was attempted for those conditions that gave well structured crystals. Fine screening conditions were selected based on the crystallization solution composition (avoiding difficult to handle compounds, such as volatile agents) and geometric structure of the crystals grown in such conditions (aiming for those crystals big enough for X-ray diffraction and with stable visual appearance). Like so, we aimed to work on crystals with high probability of being protein crystals and not salt crystals and avoided possible space groups corresponding to the ones in which uncomplexed CIN85 SH3 domains had previously crystallized.

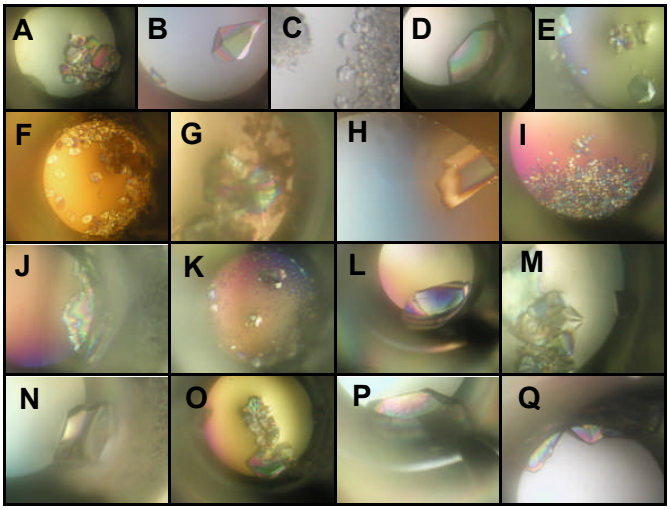


Figure V.37. Crystals of of CIN85-SH3A$_{2\text{-}58\ tga}$ obtained at 22ºC. Crystallization conditions were: (A) 2.0M $(NH_4)_2SO_4$, 0.1M Tris pH 7.0, 0.2M Li_2SO_4 ;(B) 20% (w/v) PEG-3000, 0.1M Citrate pH 5.5; (C) 1.0M Sodium Citrate, 0.1M Cacodylate pH 6.5; (D) 1.2M NaH_2PO_4/0.8 M K_2HPO_4, 0.1M CAPS pH 10.5, 0.2M Li_2SO_4; (E) 2.0M $(NH_4)_2SO_4$ CAPS pH 10.5, 0.1M Li_2SO_4; (F) 1.0M Sodium Citrate, 0.1M Imidazole pH 8.0; (G) 1.26M $(NH_4)_2SO_4$, 0.1M Tris pH 8.5, 0.2M Li_2SO_4; (H) 30% (w/v) PEG-3000, 0.1M CHES pH 9.5; (I) 1.0M K/Na Tartrate, 0.1M CHES pH 9.5, 0.2M Li_2SO_4; (J) 20% (w/v) PEG-8000, 0.1M Tris pH 8.5, 0.2M $MgCl_2$; (K) 2.0M $(NH_4)_2SO_4$, 0.1M Cacodylate pH 6.5, 0.2M NaCl; (L) 20% (w/v) PEG-3000, 0.1M Tris pH 7.0, 0.2M $Ca(OAc)_2$; (M) 1.0M Sodium Citrate, 0.1M CHES pH 9.5; (N) 1.26M $(NH_4)_2SO_4$, 0.1M HEPES pH 7.5; (O) 1.0M Sodium Citrate, 0.1M Tris pH 7.0, 0.2M NaCl, (P) 0.8M NaH_2PO_4/1.2 M K_2HPO_4, 0.1M Acetate pH 4.5; and (Q) 1.0M K/Na Tartrate, 0.1M Tris pH 7.0, 0.2M Li_2SO_4.

Crystals used for the solution of the protein complex structure were grown at 22ºC by the hanging drop vapor diffusion method in 0.8M Sodium Citrate, 0.1M Bis-Tris pH 7.5 and 0.2M NaCl (Figure V.38). Prior to flash cooling under a nitrogen gas stream, crystals were transferred to a stabilization solution of 10% (v/v) Ethylene Glycol, 0.54M Sodium Citrate, 0.09M Bis-Tris pH 7, and 0.18M NaCl.

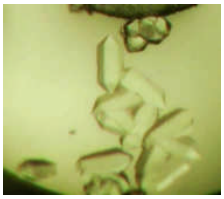


Figure V.38. Crystals of of CIN85-SH3A$_{2\text{-}58\ tga}$ used for structure determination. Crystals were obtained at 22ºC in 0.8M Sodium Citrate, 0.1M Bis-Tris pH 7.5 and 0.2M NaCl.

A 2.0Å resolution was achieved in the diffraction of the crystal and a dataset was collected at 100ºK on a rotating anode with 180 seconds exposure and a 0.75º rotation per image. Diffraction data were reduced using the HKL package, DENZO and SCALEPACK. Crystals of CIN85-SH3A$_{2\text{-}58\ tga}$:Cbl-b belong to space group $P6_5$. Unit cell parameters were as follows: $a = b = 51.158$, $c = 97.472$, $\alpha = \beta = 90º$, $\gamma = 120º$ and a completeness of 99.7%.

Initial phases for CIN85-SH3A:Cbl-b were calculated from a single AMoRe molecular replacement solution with correlation coefficient of 24.1% and R-factor of 47.6% by using the previously solved isolated CIN85-SH3A$_{2\text{-}58}$His$_6$ structure as search model. A resolution range of 15-3.5Å was used all throughout the molecular replacement process. The integration radius for the rotation function was 22.84Å, corresponding to 51% of the longest dimension of the model (model dimensions: 44.54 x 18.01 x 39.64Å). For both the rotation and translation functions the 20 orientations/translations with at least one half the highest correlation coefficient of the search value were used. The correct molecular replacement solution was found for the primitive hexagonal $P6_5$ space.

With a Mathew's coefficient of 2.34 and a calculated solvent content of 47.42% for two molecules in the asymmetric unit, a model of the solution found was generated with the program AMoRe. The correlation coefficients after rigid body refinement of the solutions from the translation search were 32.4 and 30.4, and the R values were 51.1 and 50.4%, respectively.

Analysis with SFcheck software from the CCP4 program suite allowed us to detect a *twinning* fraction of 31.3% in the cumulative intensitive distribution graph (see

Figure V.39).

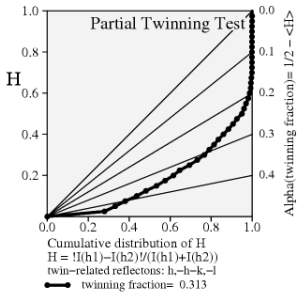


Figure V.39. Statistic analysis of the intensities of twinned data obtained with the Sfcheck tool. The experimental distribution for CIN85-SH3A:Cbl-b is plotted as a dotted line and the expected cumulative distributions of the parameter H, which is a fuction of the two twin-related intensity measurements in hemihedral twinning. Cumulative distributions for non-centrosymmetric reflections with twinning fractions of 0.0, 0.1, 0.2, 0.3, and 0.4 are represented as solid lines. The experimental distribution deviates from the theoretical for untwined data (0.0), and it resembles the most to the expected cumulative distribution of data with a twinning fraction of 0.3.

The *twinning* effect in crystallography refers to a disorder in the crystal growth that is composed of different crystal domains in different orientations but related in a particular and well defined manner. Most of the time, this change in the domain orientation does not have an effect in the crystal that can be recognized macroscopically. Usually, this *twinning* effect prevents the right determination of the structure, unless it is detected and avoided or corrected.

There are essentially two categories of *twinning*, the *epitaxial* and the *merohedral*. In the first case, *twinning* can be detected in the diffraction pattern, given that each domain in the crystal has a different diffraction pattern, and so, data from the predominant domain can sometimes be integrated by rejecting reflections of the other domain.

In the case of the *merohedral twinning*, which was our case, it refers to the special cases where crystal lattices[7] from the different domains superpose in three dimensions (Figure V.40). In our case, the unit cell of the space group to which the crystal belongs to, has two equal size sides and two equivalent angles so that these sides can be interchanged without deforming the crystal and allowing an organized crystal growth. However, the internal symmetry is changed (the asymmetric unit would not be oriented in the same way in both domains). The most common case of merohedral twinning is the hemihedral twinning, which includes only two orientations of the domains.

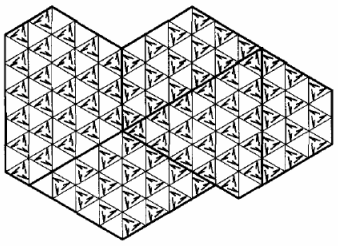


Figure V.40. "A hypothetical example of merohedral twinning in which separate domains with P3 symmetry are related by a 180° rotation about an axis parallel to the threefold crystallographic axis. The arrows represent individual molecules. The twinning is hemihedral, having two distinct domain orientations. The domain boundaries are indicated by bold lines." Taken from Yeates, T.O. (1997) (Yeates 1997).

As a result of a hemihedral twinning, the observed diffraction pattern appears to be normal, but each of the observed intensities contains contributions from two reflections that are related by the twinning operation but are crystallographically independent. In order to be able to correctly handle these data it

[7] Regular geometrical arrangement of points or particles over a 2 or 3 dimension periodic pattern.

is necessary to know the twinning fraction contained in the crystal (Yeates 1997){http://www.doe-mbi.ucla.edu/Services/Twinning/intro.html}. There are several tools that can be used nowadays for the calculation and application of this fraction for the processing of the data and solution of the structure. We used the approach of estimating the proper partitioning of each observed intensity into the two components from the separate domains. For each pair of twin-related observations the true crystallographic intensities can be calculated and recovered from the measured values (Yeates 1997).

To correct the amplitudes of the CIN85-SH3A:Cbl-b *.mtz* data file we used the Detwin program included in the CCP4 program suite applying a twinning fraction of 30% to generate a new reflection data file for the refinement cycles. We first attempted the refinement with Refmac cooperative software (Collaborative Computational Project 1994), but unsatisfactory statistics were obtained. We then proceeded to the refinement of the model with CNS (Brunger, Adams et al. 1998), with which the statistics were notably improved. The final model was obtained by combining cycles of rigid-body refinement and simulated annealing, and temperature factors in CNS considering a twinning fraction of 0.305 for the operator k, h, -l, alternating with manual model building using the programs Xfit and Coot V0.1.2-Pre-1. Electron density was detected for only one molecule of the peptide in between the two SH3 molecules. The peptide molecule was modeled manually using the program Xfit (McRee 1999). Of the total 125 residues, residue 58 in one of the SH3 molecules ($SH3A_I$) of CIN85-$SH3A_{2\text{-}58\ tga}$ was disordered, and a total of 82 water molecules were added to the final model.

Crystallographic refinement of the complexed structure against all data to 2.0Å resulted in R-factor of 22.1% and Rfree of 27.1% (for a random 5% subset of all data). This cross-validation represents acceptable values, since the Rfree/R ratio is close to unity (Tickle, Laskowski et al. 1998) (see Table V.2).

Table V.2. Crystallographic data collection and refinement statistics of CIN85-$SH3A_{2\text{-}58\ tga}$:Cbl-$b_{902\text{-}912}$ structure

CIN85-$SH3A_{2\text{-}58\ tga}$:Cbl-$b_{902\text{-}912}$ (PDB 2BZ8)	
Space group	$P6_5$
Unit Cell Dimensions (Å)	a=b=51.158, c=97.472
	α=β=90º; γ=120º
Data range (Å)	20.0-2.0
Observations (unique)	18888 (9776)
Completeness (%) (last shell)	99.7(98.9)
R_{sym} [a] (last shell)	0.062 (0.430)
Reflections F>0 (cross validation)	9742 (746)
Non-hydrogen atoms (solvent molecules)	1027(82)
R_{cryst} [b] (R_{free} [c])	22.1. (27.1)
R.m.s. bond length (Å)	0.007
R.m.s. bond angles (º)	1.334

[a] R_{sym} is the unweighted R value on I between symmetry mates.
[b] $R_{cryst} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$
[c] R_{cryst} is the crossvalidation R factor for 5% of reflections against which the model was not refined.

Unexpectedly, we observe a structure of a ternary complex that is composed of two SH3 domains and a single Cbl-b peptide, as mentioned previouly. Both CIN85-SH3A domains adopt the typical SH3 domain fold, a five stranded β-sandwich, and overlap with one another with r.m.s. deviations between Cα atoms of 0.5Å. In this heterotrimeric complex, two SH3 domains contact the Cbl-b peptide, which itself adopts a classical PPII helical conformation, through their conserved ligand-binding surfaces (Figure V.41). However, one SH3 engages the peptide resembling a class I interaction (defined as SH3 I), whereas the other binds on the opposite side of the peptide ligand resembling a class II orientation (defined as SH3 II). Recognition of the peptide does not correspond to a canonical class I or class II SH3-peptide binding in either case, since none of them binds the prolines from the motif in the same way required for a canonical SH3:PPII binding. This is such that the RT loop from SH3 I recognizes the N-terminal arginine in the Cbl-b polyproline motif, and SH3 II does so with an arginine in the Cbl-b C-terminus, but none of the SH3 molecules recognize proline residues with their hydrophobic pockets in the canonical mode.

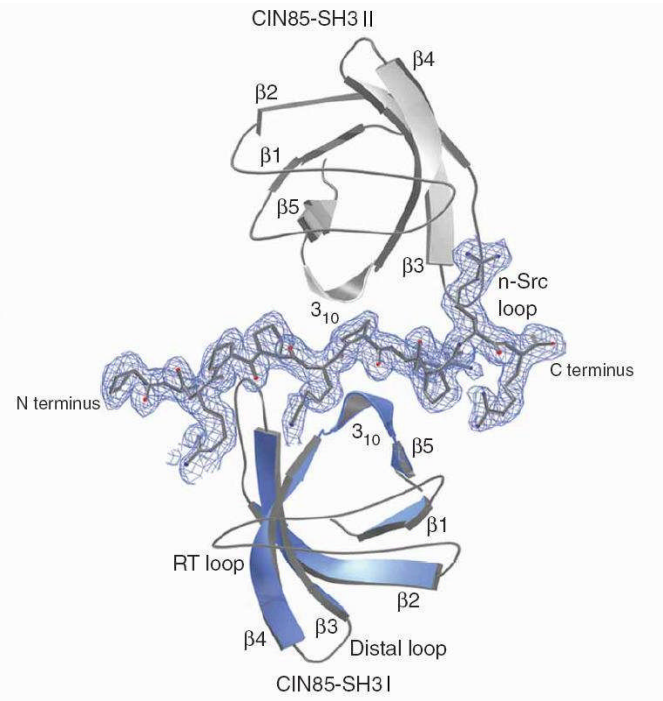


Figure V.41. Overall structure of the CIN85-SH3A:Cbl-b heterotrimeric complex. The electron density map around the peptide is contoured at 2.5σ. The secondary structural elements and position of the RT, n-Src and distal loops are indicated. The SH3A that binds the peptide in a class I-like orientation is colored in blue, while the SH3A that recognizes the Cbl-b peptide in a class II-like orientation is shown in grey.

Therefore, the recognition of an arginine in each extreme of the Cbl-b peptide by the acidic pocket of each SH3 domain provides a different recognition orientation to each SH3. As seen in Figure V.42, Arg909 interacts with both domains simultaneously, occupying the position of a conserved proline in both hydrophobic

pockets I and II from SH3 I and SH3 II respectively. The other hydrophobic pockets establish typical hydrophobic contacts that have been described for canonical binding. Therefore, we refer to the binding of SH3 I as class I-like binding, and class II-like binding to the peptide recognition of SH3 II.

The peptide is thus sandwiched by the two SH3 domains in the ternary complex and there are no significant direct SH3:SH3 contacts in the CIN85-SH3A:Cbl-b structure. Only aspartic acid from the n-Src loop of each SH3 domain (Asp33) and Arg54 establish Van der Wals contacts, that don't interfere in the peptide binding and this indicates that the heterotrimer formation is mediated almost exclusively by the peptide binding.

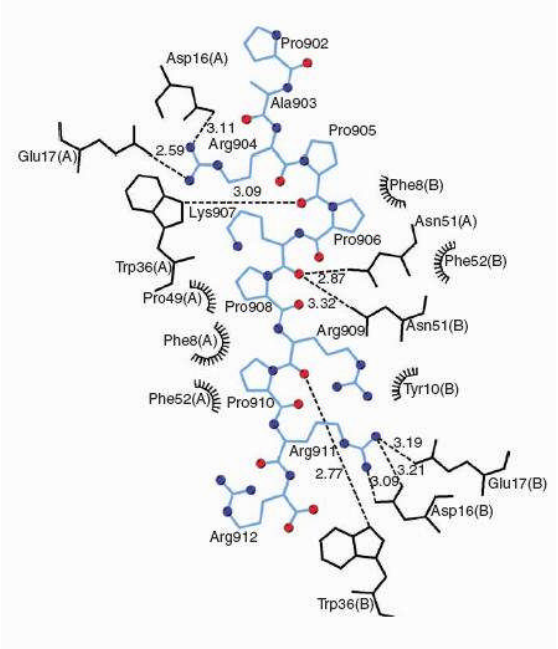


Figure V.42. Schematic representation of the protein-peptide contacts between the Cbl peptide and CIN85-SH3A. The Cbl-b peptide is represented in blue, and residues from CIN85-SH3A in black. Black dashed lines denote hydrogen bonds and black rays, hydrophobic interactions. The SH3 domain which the residue belongs to is indicated in parentheses: A stands for class I-like SH3 and B for class II-like SH3.

Complex formation between the Cbl-b peptide and either SH3 domain buries 928Å^2 (SH3 I) or 715Å^2 (SH3 II). These values are comparable to those of other SH3 domain:peptide structures. Specific interactions via peptide side chains are only made by Arg904 and Arg911 from Cbl-b at the extreme N and C termini of the peptide with Asp16 and Glu17 of each RT loop of each CIN85-SH3A molecule (see Figure V.42). Argines 904 and 911 are thus responsible for the binding orientation of each SH3 domain in the ternary complex.

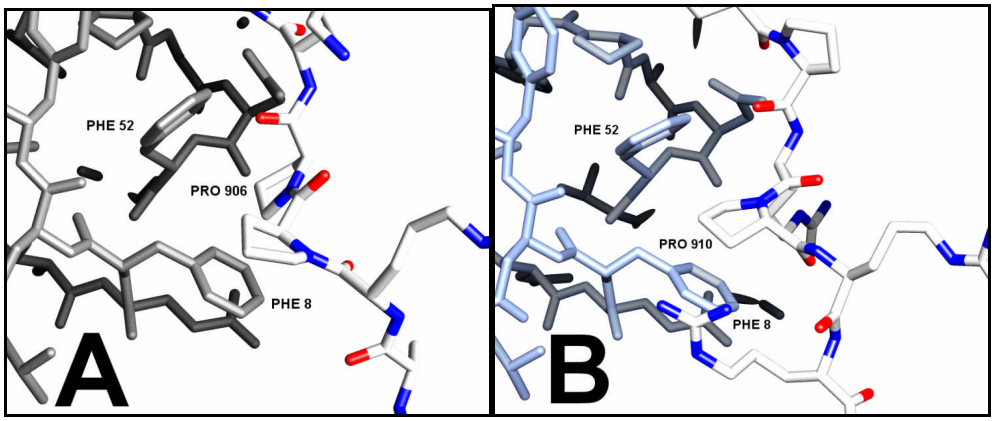


Figure V.43. Detail of the CIN85-SH3A:Cbl-b heterotrimeric structure. Prolines 906 (A) and 910 (B) from the Cbl-b peptide make hydrophobic contacts with phenylalanines 8 and 52 in SH3 II (*left*) and SH3 I (*right*). The peptide is shown in white with oxygen atoms in red and nitrogen atoms in blue, and SH3 I and II are shown in blue and grey, respectively.

Hydrophobic contacts are made by Pro906 and Pro910 from Cbl-b with conserved phenyalalnine residues 8 and 52 from hydrophobic pocket I of each SH3 domain (SH3 II and SH3 I, respectively; see Figure V.43), whereas Pro908 is surrounded by hydrophobic residues from hydrophobic pocket II of SH3 II in the CIN85-SH3A:Cbl-b complex (Tyr10, Trp36, Pro49 and Phe52; see Figure V.44).

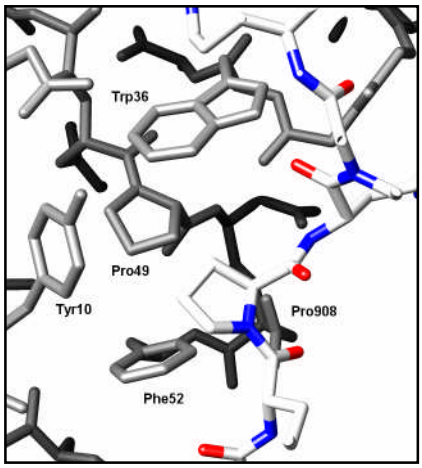


Figure V.44. Detail of the hydrophobic pocket II in SH3 II of the CIN85-SH3A:Cbl-b heterotrimeric structure. Proline 908 from the Cbl-b peptide makes hydrophobic contacts with Tyr10, Trp36, Pro49 and Phe52 from CIN85-SH3A$_{SH3\ II}$. The peptide is shown in white with oxygen atoms in red and nitrogen atoms in blue, and SH3 II is represented with grey sticks.

Several hydrogen bonds are formed by backbone carbonyl oxygens of the peptide to side chains of either SH3 domain. The indole rings from Trp36 from each SH3 domain hydrogen bonds the carbonyl oxygen from Cbl-b Pro905 in the class I-like recognition, and the carbonyl oxygen from Arg909 in the class II-like orientation (see Figure V.42).

Interestingly, Lys907 from the Cbl peptide plays a central role since it is the only residue that makes simultaneous contacts with both SH3 domains through hydrogen bonds from its backbone carbonyl to the side chain amides of each Asn51 in CIN85-SH3A (see Figure V.42). These contacts are pseudo-symmetrical and the backbone carbonyl of Lys907 constitutes the pseudo dyad axis of these interactions. Accordingly, equivalent interactions are established by Arg904 and Arg911 as well as Pro905 and Arg909 from Cbl-b (Figure V.45).

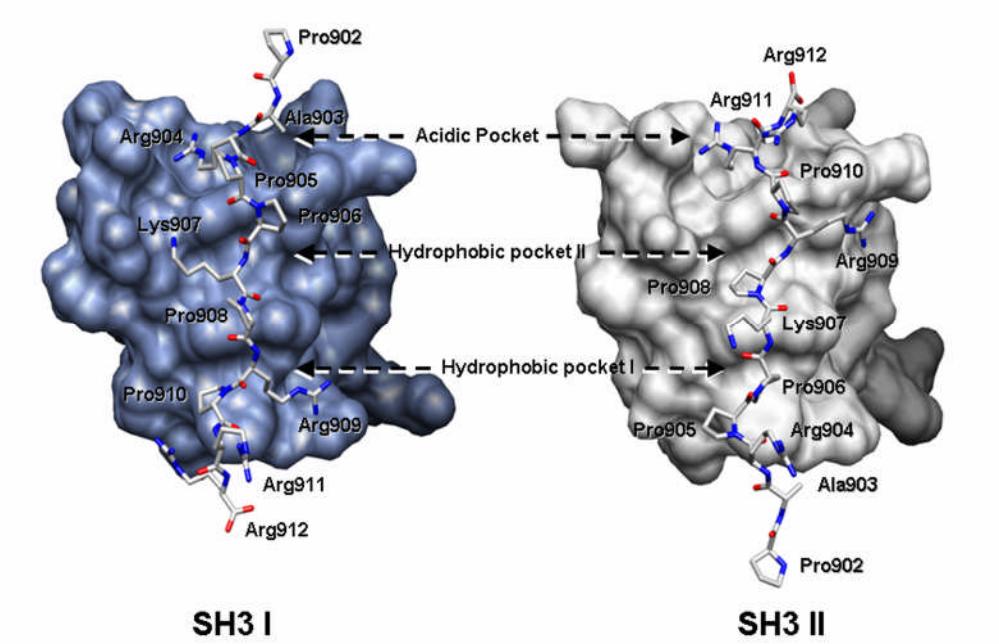


Figure V.45. Comparison of the SH3 I:Cbl-b and SH3 II:Cbl-b interfaces from the CIN85-SH3A:Cbl-b heterotrimer complex. Surfaces of CIN85-SH3A domains engaging the Cbl-b peptide in a class I-like orientation (blue, *left*) and in a class II-like orientation (grey, *right*) are shown. The Cbl-b peptide is represented in sticks, and its residues are indicated. Binding pockets are indicated with arrows.

We therefore observe that essential residues in the interaction appear to mantain a symmmetrical distribution. Key to the CIN85-SH3A recognition are Cbl-b arginines 904 and 911 that are recognized by the RT loops of each SH3 domain, and which are determinant for the SH3 orientation in this novel heterotrimer arrangement.

Validation of the quality of the model of the CIN85-SH3A:Cbl-b complex was undertaken with the program PROCHECK from the CCP4 program suite and web-based tool MolProbity (Lovell, Davis et al. 2003). Analysis in both programs indicated that 98.3% of the residues were in the allowed regions of Φ and ψ conformational space, with Glu32 and Asn42 from SH3 II as outliers (see Figure V.46). These residues are located in the n-Src loop, which shows alternate conformation probably due to its flexible nature. All the other geometrical parameters were considered to have acceptable values.

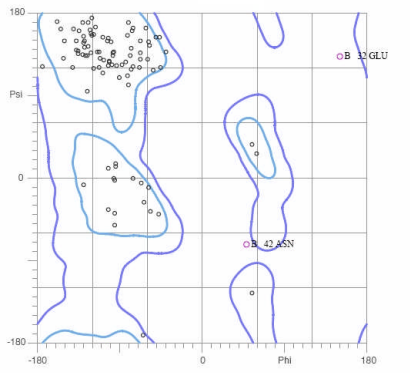


Figure V.46. Ramachandran plot of CIN85-SH3A:Cbl-b structure validation in MolPrpb. Asparagine 42 in the flexible n-Src loop from one of the SH3A molecules (SH3A $_{II}$, chain B) is in a disallowed region, and Glu32 in the same SH3 and Asn42 from the other SH3 molecule are in the generously allowed regions. All the other residues (97%) are in the favored and allowed regions.

V.1.1.3.1.1. Crystallization and structure of the CIN85-SH3A$_{2-58}$:CD2$_{324-333}$ complex

All CIN85/CMS SH3 domains can recognize the only atypical PxxxPR proline-arginine motif present in the C-terminal region of Cbl and Cbl-b (Kowanetz, Szymkiewicz et al. 2003; Kurakin, Wu et al. 2003). Additional cellular proteins, such as CD2 receptor (CD2), BLNK, AIP1/Alix or ZO2 contain this consensus sequence and have been reported to be CMS/CIN85 natural partners in mammalian cells (Dikic 2002; Kowanetz, Szymkiewicz et al. 2003). However, except for ZO2, they all lack the N-terminal arginine responsible for class I-like recognition of Cbl-b by CIN85-SH3A (Arg904$_{Cbl-b}$), and hence, class I-like recognition of the peptide in the interaction of the N-terminal SH3 of CIN85 with these partners would not happen in the same manner as we observed in the CIN85-SH3A:Cbl-b complex.

Analysis of the CIN85-SH3A:Cbl-b structure has identified key component residues in the peptide that might help predict binding of other peptides. The study of the molecular recognition of the CIN85-SH3A domain and other binding partners might shed light to the differences of the multiprotein complex formation and their functional biological consequences in the cell. Therefore, we wanted to explore the molecular interaction of the SH3A domains from CIN85 and a CD2 derived peptide containing the already identified PxxxPR motif, which differs from the recently extended CIN85 binding motif in Cbl.

Both N-terminal CIN85 and CMS homologue SH3A domains have a stronger affinity for CD2 with similar dissociation constants (K_D≈100μM) than the rest of CIN85 and CMS SH3 domains (Dustin, Olszowy et al. 1998; Hutchings, Clarkson et al. 2003).

The N-terminal SH3 domain from CIN85 was crystallized in complex with a CD2 derived peptide in order to compare its structure with the one previously solved in the laboratory of CMS-SH3A:CD2 complex. It would also help us analyze the effect of the absence of the N-terminal arginine of the Cbl motif on the molecular recognition of CIN85.

Crystals were obtained with a 1:3 purified CIN85-SH3A$_{2-58\ tga}$ protein:CD2$_{324-333}$ peptide mix, by the sitting drop vapor diffusion method in 200:100nl drops (protein solution:reservoir solution) set with a Cartesian liquid handling robot, and

grown at 22ºC. Very big and twinned crystals were obtained in three conditions (see Figure V.47), so surgery had to be practiced on them to be able to separate manageable sized single crystals feasible for crystal diffraction. Best diffraction was measured with a crystal grown in 0.1M CHES pH 9.5, 1M Na Citrate in the reservoir solution.

Figure V.47. Crystals of CIN85-SH3A$_{2\text{-}58\ tga}$:CD2$_{324\text{-}333}$ complex. Crystals were grown by the sitting drop vapor diffusion method in (A) 30% (w/v) PEG-3000, 0.1M CHES pH 9.5, (B) 2.0M $(NH_4)_2SO_4$, 0.1M Cacodylate pH 6.5, 0.2M NaCl and (C) 1M Na Citrate, 0.1M CHES pH 9.5.

Crystals were flash cooled in the reservoir solution, since it did not form ice and no additive was needed in the reservoir solution for cryo protection. They were stored in a liquid nitrogen container to be diffracted in a synchrotron radiation source, at the ID23-1 beamline at the ESRF facility in Grenoble, France. Initial diffraction pattern showed resolution beyond 1.0Å, but the rapid crystal decay only allowed us to collect a final dataset up to 1.2Å resolution to ensure good quality data. A dataset of 180 images of 1º rotation in phi, with 0.05 seconds exposure was collected (Figure V.48).

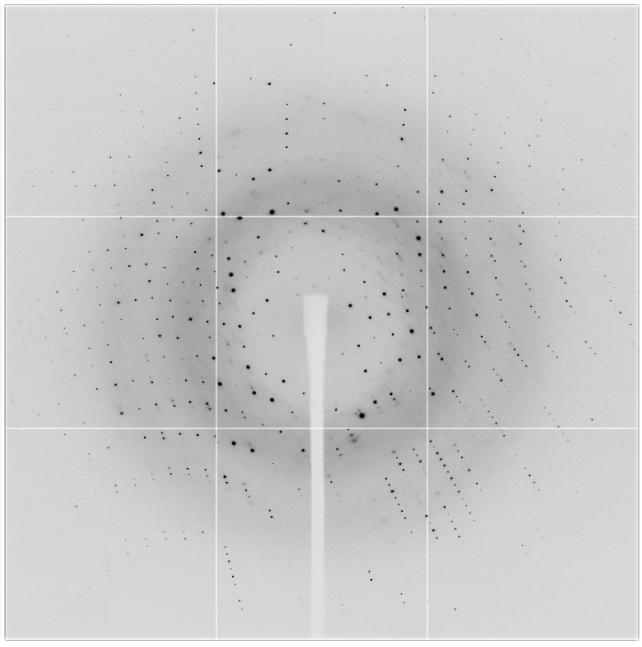
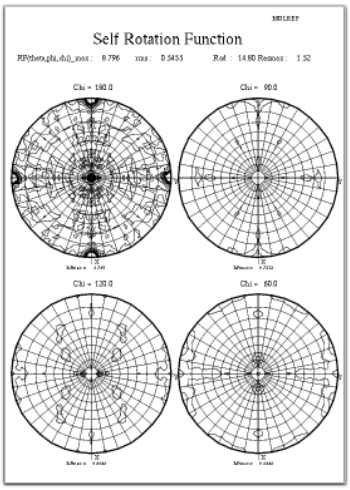


Figure V.48. Diffraction Image from the crystal of the complex of CIN85-SH3A$_{2\text{-}58\ tga}$:CD2$_{324\text{-}333}$. This image corresponds to a diffraction exposure of 0.05 seconds in a 1º oscillation in phi, with maximum resolution at 1.2Å.



Data reduction was carried out with Scala software. Tails correction with width of 0.1 and a slope of 10 was applied and standard deviation correction values of 0.045 for full reflections and 0.035 for partials. Self rotation Function showed peaks for three binary axes (see Figure V.49), indicating that the data were of a P222 space group. Systematic absences were detected every two reflections in the k and l axes. Accordingly, data was re-indexed with an "l, -k, h" transformation matrix, following the IUCr (International Union of Crystallography) convention (Figure V.50).

Figure V.49. Self Rotation function of data from CIN85-SH3A:CD2 crystal. Peaks corresponding to three binary axes (one perpendicular to the plane, and two on the plane), all related by 90º rotation, can be observed.

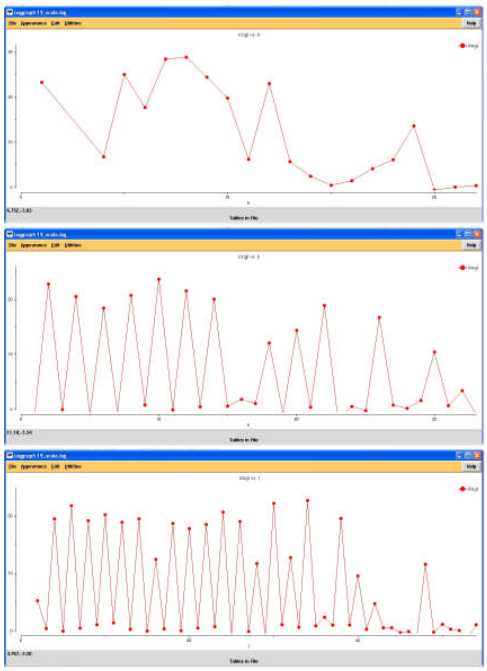


Figure V.50. Plotted I/σ vs each reflection axis: (A) h, (B) k, (C) l. Systematic absences are detected every two reflections in the k and l axes. This condition is not contemplated in any orthogonal space group condition. In order to fullfil the IUCr convention of P222 space group, a transformation with an l, -k, h matrix should be applied, so that systematic absences are in the h and k axes.

Structure was solved by molecular replacement using Phaser (ccp4), and a pdb file containing only an SH3 domain from the CIN85-SH3A:Cbl-b complex (PDB 2BZ8, chain B). Then, automated model building was carried out with ARP/wARP. A very defined electron density area was already observable in this model, where the peptide should bind (see Figure V.50). After a quick refinement with Refmac5, superposition of the CMS-SH3A:CD2 dimeric complex (PDB 2J7I), previously solved in the laboratory by Gabriel Moncalián, allowed the positioning of the CD2 peptide in the CIN85 model (see Figure V.53). The peptide did not adopt the same

conformation as in the CMS-SH3A:CD2 model, which was corrected after some cycles of manual (with Coot v0.1.2-pre-1) and automatic refinement (with Refmac5).

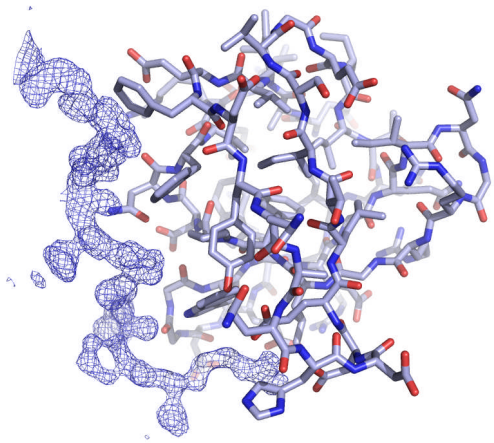


Figure V.51. Preliminary model of the CIN85-SH3A:CD2 complex after automatic rebuilding in ARP/wARP. A difference map contoured at 1.0σ around the peptide is shown in a blue mesh, where residues from the CD2 peptide can readily be deduced (ie. Arg332 and Pro331 side chains become evident at the bottom of the picture). The CIN85-SH3A molecule is shown in sticks.

Extensive refinement was carried out in Coot and Refmac5 until R and R-free factors reached values of 20.8% and 20.9%, respectively. Pdb and mtz files were transformed into ins and hkl files (the latter conserving the same Free-flags) using ShelxPro to continue the refinement in Shelxl.

First refinement in Shelxl (Sheldrick, Schneider et al. 1997) was performed without the addition of water molecules and only up to 2.5Å, resolution was then increased to 1.5Å and finally to 1.2Å.

Due to the differences in the bulk solvent model used in SHELXL, the difference between the R and the R-free values increased, compared to the values obtained in Refmac5 (Sheldrick, Schneider et al. 1997) (Table V.3).

Table V.3. Crystallographic data collection and analysis of CIN85-SH3A$_{2\text{-}58\ tga}$:CD2$_{324\text{-}333}$ structure

CIN85-SH3A$_{2\text{-}58\ tga}$:CD2$_{324\text{-}333}$	
Space group	$P2_12_12_1$
Unit Cell Dimensions (Å)	a=26.9, b=39.6, c=65.5
	α=β=γ=90º
Data range (Å)	39.6-1.2
Observations (unique)	153684 (22657)
Completeness (%) (last shell)	100 (99.8)
R_{sym} [a] (last shell)	0.076 (0.652)
Reflections F>0 (cross validation)	21498 (1159)
Non-hydrogen atoms (solvent molecules)	657 (89)
R_{cryst} [b] (R_{free} [c])	0.207 (0.228)
R.m.s. bond length (Å)	0.010
R.m.s. bond angles (º)	1.391

[a] R_{sym} is the unweighted R value on I between symmetry mates.
[b] $R_{cryst} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$
[c] R_{cryst} is the crossvalidation R factor for 5.1% of reflections against which the model was not refined.

Validation of the geometric parameters in MolProbity (Lovell, Davis et al. 2003) showed all residues in the favored residues with Asn42 as an outlier, which is highly exposed to the solvent (Figure V.52). Lys58 and Gln11 were identified as rotamer outliers differing slightly from the standard values.

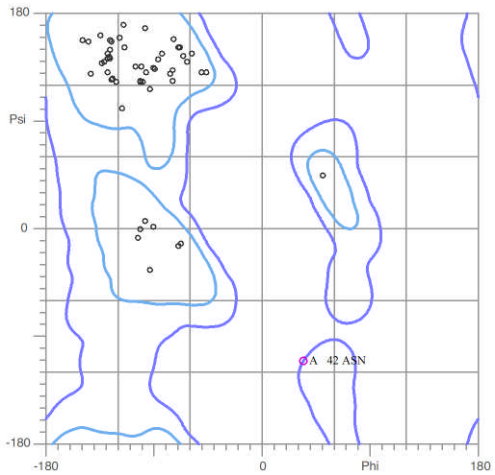


Figure V.52. Ramachandran plot of CIN85-SH3A:CD2 structure validation in MolProbity webtool. 98.6% of the residues are in the favored regions, with only Asn42 from the n-Src loop as an outlier.

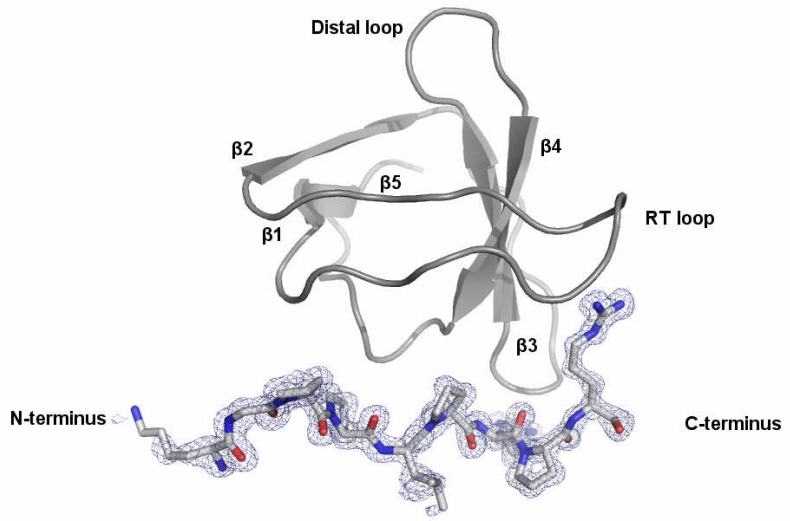


Figure V.53. Overall structure of the CIN85-SH3A:CD2 heterodimeric complex. The electron density map around the peptide is contoured at 1.0σ. The secondary structural elements are indicated. CIN85-SH3A, which binds the CD2 peptide in a class II-like orientation is colored in grey.

Not surprisingly, the structure of the complex showed a 1:1 recognition of the peptide, where the SH3 recognizes the CD2 peptide in a class II-like recognition mode (Figure V.53). The first proline in the motif, Pro327, occupies the hydrophobic pocket II, and Arg330 occupies the position in hydrophobic pocket I, usually filled by a proline in a classical class II recognition (Figure V.54).

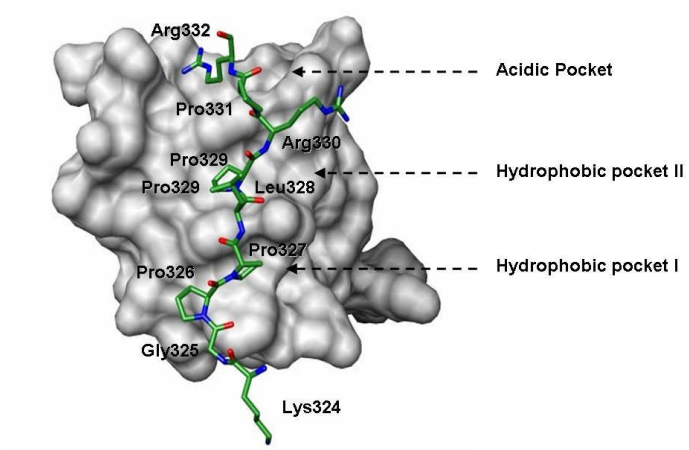


Figure V.54. CIN85-SH3A:CD2 interface. Surface of CIN85-SH3A domain is colored in grey, and the CD2 peptide is shown in green sticks. Peptide residues are indicated and SH3 binding pockets are pointed out with arrows.

Arginine 332 in the C-terminal end of the CD2 peptide hydrogen bonds Asp16 and Glu17 in the CIN85-SH3A RT loop, similarly to Cbl-b Arg911 in the CIN85-SH3A:Cbl-b complex. Hydrogen bonds with carboxylic oxygens from the

peptide main chain are found between Arg330 and the Trp36 indole ring, and Leu328 and Asn51 side chain, as Cbl-b Lys907 and Pro910, respectively, in the SH3 II from the CIN85-SH3A:Cbl-b structure. Hydrophobic interactions described in the hydrophobic pocket II from the class II-like SH3 domain in the heterotrimeric structure are found as well with Pro329 in the dimeric complex with the CD2 peptide. CD2 proline 327 also establishes hydrophobic interactions with CIN85 phenylalanines 8 and 52.

We therefore observe that recognition of the CD2 peptide by CIN85-SH3A is equivalent to class II-like recognition of the Cbl-b peptide. Such that interacting residues CD2 Arg332, Pro329 and Pro327 play the same role as Cbl-b Arg911, Pro908 and Pro906 in the class II-like interaction in the heterotrimer complex of CIN85-SH3A:Cbl-b complex.

2. CMS

CIN85 and CMS are homologous proteins that have a similar domain distribution in which the main difference is the presence of several putative actin binding sites in CMS, that are absent in CIN85. They share 40% sequence identity (according to sequence alignment with ClustalW), and conserved residues are contained in the SH3 domains, in the proline-rich region and in the CC domain, which has led to the assumption that these proteins are members of the same family and share similar functions.

However, they display a differential expression pattern. CIN85 have been reported to be ubiquiotiously expressed, mainly in brain, whereas the mouse ortholog of CMS, CD2AP, isn't expressed in either brain, kidney nor lung (Kirsch, Georgescu et al. 1999; Take, Watanabe et al. 2000). They also seem to have different affinities for their different partners. It has been observed that c-Cbl binds preferentially to SH3B in CMS and binding to its SH3C is not detected. However, all SH3 domains from CIN85 bind c-Cbl, being the N-terminal domain the preferred one (see Figure V.55).

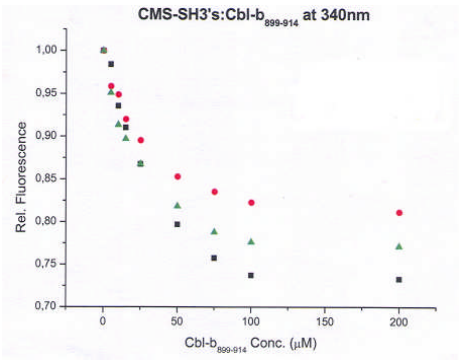


Figure V.55. Binding of c-Cbl peptide to CMS-SH3A and CMS-SH3B monitored by collisional quenching of CMS SH3 fluorescence. The change in the fluorescence intensity of CMS-SH3A upon c-Cbl binding (■) is larger than that of CMS-SH3B (●). An equimolar mixture of CMS SH3A and SH3B (▲) showed and average of SH3A and SH3B binding to c-Cbl.

In order to understand the molecular details of the different members of the CIN85/CMS family of adaptor molecules we wanted to compare the interaction of these two adaptor molecules with Cbl.

2.1. The N-terminal SH3 domain of CMS is the closest to CIN85-SH3A

Alignment of CIN85-SH3A and CMS-SH3A domain sequences shows a 68% identity, and 20% strong similarity (according to Clustal alignment results; see alignment in Figure V.56). SH3 domains B and C share 68% identity with 14% strong similarity and 54% identity with 18% strong similarity, respectively. Hence, SH3A was the best candidate for comparison of complexes.

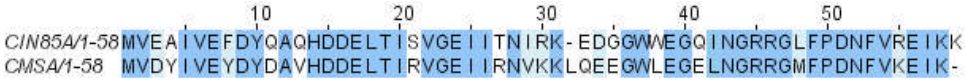


Figure V.56. Sequence alignment of the N-terminal SH3 domains from CMS and CIN85. Identical residues are shown in *blue* and conserved hydrophobic residues in *light blue*.

2.1.1. CMS-SH3A:Cbl-b complex

2.1.1.1. CMS-SH3A:Cbl-$b_{902-912}$ structrure refinement

The structure of the CMS-SH3A domain in complex with the short Cbl-b peptide was solved in the laboratory by Gabriel Moncalián at the end of his postdoctoral stay and could not be completed.

Additional refinement was carried out with cycles of manual refinement in Coot v0.1.2-Pre-1 and automatic refinement in Refmac5. In order to do so, the .cv structure factor file used for refinement in CNS had to be transformed to an mtz file, being very cautious on conserving the Free-flags from the .cv file. This way, we would be able to continue the refinement from the last CNS model with unbiased statistics. Refinement converged with 30 cycles using Babinet scaling and 10 previous TLS cycles with a TLS model containing only the SH3 domain and setting the initial temperature factors to 25.

Final R and R-free values were improved to 19% and 21.8% respectively (see Table V.4), and no significant geometry problems were detected under the Procheck and the Whatcheck tests. All residues (68/68) were located in favored regions of the Ramachandran plot (see Figure V.57). Other geometric parameters showed acceptable values. Structure files were deposited in the Protein Data Bank database with PDB code 2J6F.

Table V.4. Crystallographic data collection and analysis

CMS-SH3A:Cbl-b$_{902-912}$ (PDB 2J6F)	
Space group	I422
Unit Cell Dimensions (Å)	a=66.51, b=66.51, c=67.78
	α=β=γ=90º
Data range (Å)	20.0-1.7
Observations (unique)	205566 (8696)
Completeness (%) (last shell)	100.0 (99.9)
R_{sym} [a] (last shell)	0.034 (0.269)
Reflections F>0 (cross validation)	8534 (465)
Non-hydrogen atoms (solvent molecules)	547 (49)
R_{cryst} [b] (R_{free} [c])	0.243 (0.277)
R.m.s. bond length (Å)	0.0055
R.m.s. bond angles (º)	1.58

[a] R_{sym} is the unweighted R value on I between symmetry mates.
[b] $R_{cryst} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$
[c] R_{cryst} is the crossvalidation R factor for 5% of reflections against which the model was not refined.

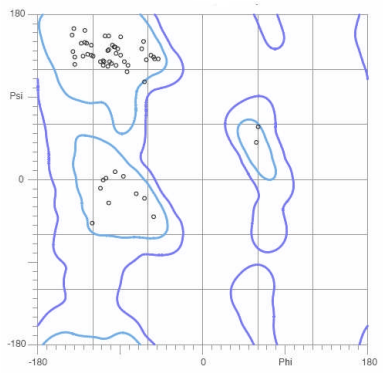


Figure V.57. Ramachandran plot of CMS-SH3A:Cbl-b structure complex validation in MolProb program. There are no residues in disallowed regions and 96.2% of them are in the favored regions and 2 residues in the allowed regions.

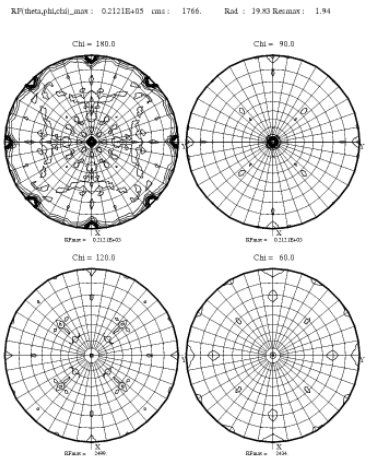


Figure V.58. Self Rotation function of data from CMS-SH3A:Cbl-b crystal. Peaks corresponding to two binary axes and a quaternary axis can be observed.

This crystal showed an I422 symmetry (see rotation function in Figure V.58), and the structure of the complex displays one CMS-SH3A molecule bound to a single Cbl-b peptide in the asymmetric unit. However, we observed that the peptide lies perpendicular to a crystallographic 2-fold axis. Thus, electron density maps show an average of two orientations for the peptide (Figure V.59). Final refined B factors for the peptide, considering occupancy of 0.5 for each orientation, do not differ significantly from the B factors of the CMS-SH3A interacting residues (~25Å^2). The double orientation of the peptide is possible, because crystallographic contacts

are exclusively established by CMS-SH3A molecules. The model could not be constructed for CMS-SH3A residues 1 and 59–62, since they were disordered and thus not visible in the electron density maps. Cbl-b residues 902, 903, and 904 were also disordered, meaning that they did not participate directly in the interaction with the CMS-SH3A domain.

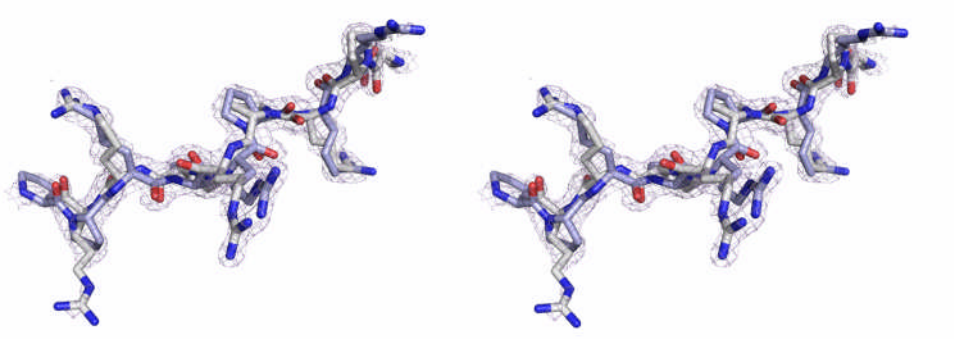


Figure V.59. Stereo view of symmetrically related peptides from the CMSA:Cbl-b heterotrimeric complex shows the double orientation of the peptide. In this representation one of the symmetrically related peptides is colored in grey and the other one is colored in light blue. SigmaA weighted electron density map around them is contoured at 1.0σ.

A deeper anlalysis of the structure showed us that a heterotrimer can be formed by symmetry. As we observed in the CIN85-SH3A:Cbl-b complex, the structure of the CMS-SH3A:Cbl-b complex consists on a ternary complex composed of two CMS-SH3A molecules binding, in opposite orientations, to a single Cbl-b peptide molecule (Figure V.60). However, in contrast to CIN85-SH3A:Cbl-b complex, one CMS SH3A domain recognizes Lys907 in the N-terminus of the peptide (as opposed to Arg904 in the case of CIN85-SH3A), similarly resembling a classical class I orientation, and the second SH3 domain recognizes Arg911 in the C terminus, resembling class II binding (the same way that it was described for the CIN85-SH3A:Cbl-b complex; Figure V.61).

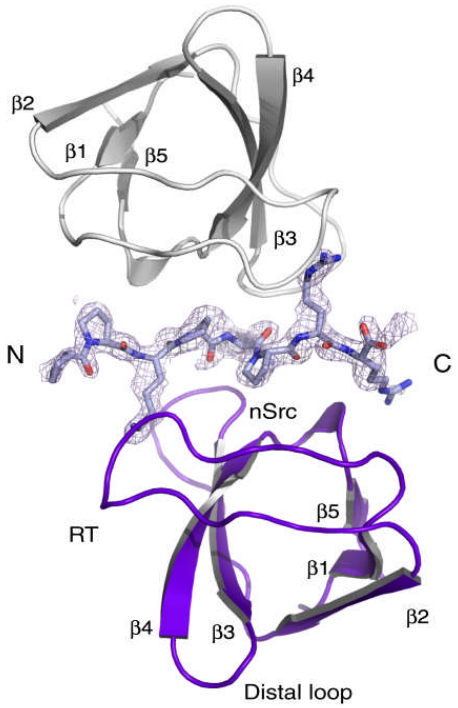


Figure V.60. Overall structure of the CMS-SH3A:Cbl-b heterotrimeric complex. CMS-SH3A$_{SH3\ I}$ binds the peptide in a class I-like orientation (shown in purple), and CMS-SH3A$_{SH3\ II}$ recognizes the peptide in class II-like orientation (shown in gray). The Cbl-b peptide is shown in light blue. The σA-weighted electron density map around it is contoured at 1.0σ. Elements of secondary structure and the positions of the RT, n-Src, and distal loops are indicated.

The structure of the CMS-SH3A molecule also displays the typical SH3 domain folding of a five-stranded β-barrel. Such β-barrel is stabilized by the hydrophobic nucleus, which is formed by the same residues as in CIN85-SH3A, except that the conserved tryptophan (CIN85 Trp37) is a leucine in CMS (CMS Leu38). The peptide, on the other hand, does not display a typical polyproline helix but adopts an extended conformation instead.

The structure shows that CMS also has the ability to form heterotrimeric complexes with a Cbl-b derived peptide, as we observed in the CIN85-SH3A complex. In the heterotrimer arrangement, the interactions of theCMS-SH3A$_{SH3\ I}$ (class I-like orientation) and CMS-SH3A$_{SH3\ II}$ (class II-like orientation) with the Cbl-b peptide are almost identical with 941 and 827Å^2 of buried surface, respectively. However, the interacting surface of the two CMS-SH3A molecules is negligible. Only 54Å^2 of the 3918Å^2 total surface of each SH3 is shared, corresponding mainly to side chains of Glu35 (from both CMS-SH3A$_{SH3\ I}$ and CMS-SH3A$_{SH3\ II}$).

In this case, hydrophilic specific interactions via side chains are only made by Lys907 and Arg911 at the N and C termini of the Cbl-b peptide with both Asp16 and Glu17 from CMS-SH3A$_{SH3\ I}$ andCMS-SH3A$_{SH3\ II}$, respectively and Arg909 with Glu35 from both CMS-SH3A molecules. Hydrophobic contacts in the class I-like orientation are made by Cbl-b Pro910 with conserved residues of hydrophobic pocket II. Hydrophobic pocket I in the class I-like binding is empty due to the limited extension of the peptide. In turn, Cbl-b Pro906 and 908 are embedded within hydrophobic pockets I and II, respectively in the class II-like orientation. Thus, Cbl-b Pro908 and 910 occupy equivalent positions in the opposed SH3 domains of the heterotrimer, as observed in the CIN85-SH3A:Cbl-b heterotrimer.

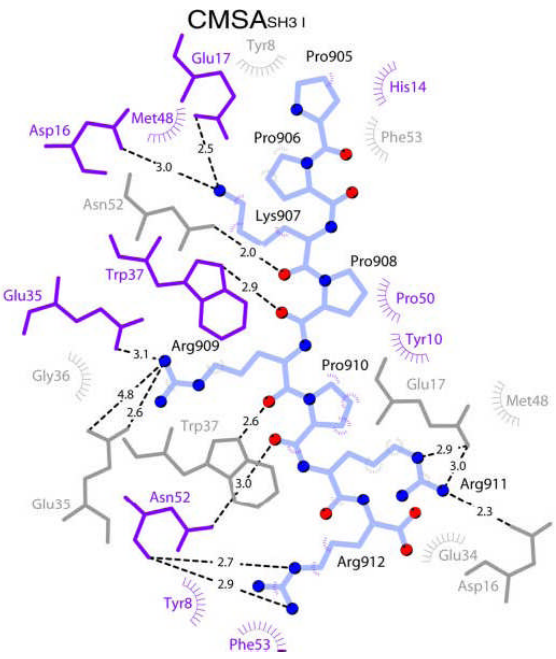


Figure V.61. Schematic representation of contacts between CMS-SH3A and the Cbl-b peptide (light blue). Residues 902, 903, and 904 from the Cbl-b peptide were disorderer, and hence not visible at the electron density map. Interactions and residues from SH3 I are shown in purple and in gray for SH3 II. Dashed lines show hydrogen bonds (labeled with peptide-protein distances in Å), and purple and gray rays designate hydrophobic interactions.

Additionally, there are sequence-unspecific contacts with peptide main chain atoms, such as Pro910 and Lys907 carboxylic oxygens interacting, respectively, with Asn52 from each CMS-SH3A$_{SH3\ I}$ and CMS-SH3A$_{SH3\ II}$. Similarly, Pro908 and Arg909 carboxylic oxygens interact with Trp37 from each CMS-SH3A.

2.1.1.2. CMS-SH3A:YCbl-b$_{903-914}$ structure

The crystal structure of the CMS-SH3A:Cbl-b$_{902-912}$ complex called our attention on the difference in the class I-like recognition of the Cbl-b peptide in comparison with the CIN85 complex. The different orientation of the SH3 domains did not allow any peptide residue to interact with the hydrophobic pocket I from the SH3 I. Therefore, we aimed to study the full interaction of the SH3A from CMS with a longer Cbl-b peptide extended in the C-terminal end.

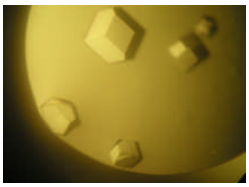


Figure V.62. Crystals of the CMS-SH3A:YCbl-b$_{903-914}$. These crystals were grown at 20ºC with 35% PEG 3k, 0.1M Tris pH 6.5, 0.2M NaCl reservoir solution.

We crystallized the N-terminal SH3 domain of CMS in complex with the Cbl-b derived peptide, two residues longer in the C-terminal end than the one used for the crystallization of the CIN85-SH3A:Cbl-b$_{902-912}$ and CMS-SH3A:Cbl-b$_{902-912}$ complexes. As it was the case for CIN85-SH3A, CMS-SH3A was expressed untagged and purified by ammonium sulfate precipitation. Crystals were obtained at 20ºC by hanging vapor diffusion method, with 1µl drops of a 1:1 protein mix at 8.1mM (60mg/ml) concentration and reservoir solution with 30% PEG 3kD, 0.1M Tris pH 7, 0.2M NaCl (Figure V.62). Crystals diffracted to a 1.65Å resolution and belonged to the I422 space group with a unit cell similar to the one calculated for the previous CMS-SH3A:Cbl-b model, defined by a = b = 66.47, c = 68.06 and angles α = β = γ = 90º (see Table V.5).

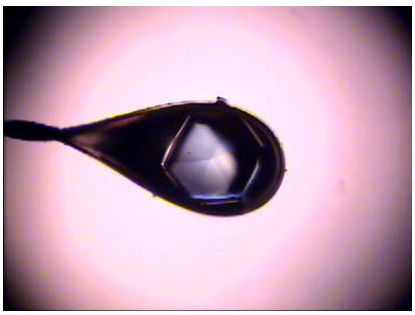


Figure V.63. Crystal of CMS-SH3A:YCbl-b$_{903-914}$ complex, diffracted in the synchrotron radiation source at the ESRF facility at beamline ID-14.

Two datasets were collected with a 0.934Å beam in the synchrotron radiation source at the ESRF facility from a single crystal (Figure V.63), one with a high resolution limit of 1.4Å and a second one to ensure complete collection of low resolution diffraction (Figure V.64). Each image was of a 1º rotation with 1 second of exposure. Data reduction was undertaken with Mosflm software, and scaled limiting the resolution to 1.65Å for better statistics, with isotropic B-factors and standard

correction protocol with SDadds of 0.03 for full reflections and 0.02 for partial reflections (Scala, ccp4). A 99.9% completeness (with a maximum of 8 overloads and bad spots in two images) was achieved with the processing of the high resolution data, so that merging with the low resolution data was not necessary.

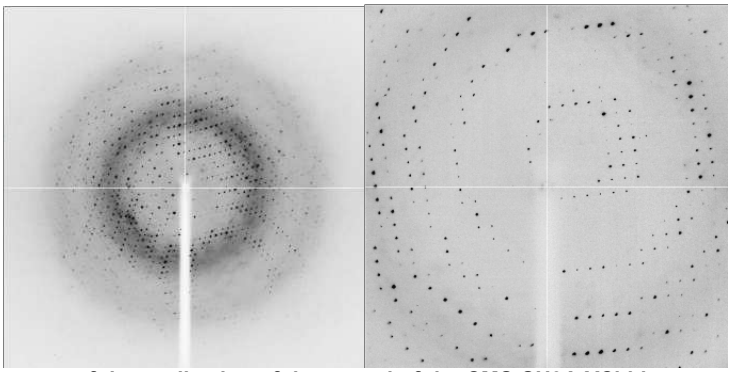


Figure V.64. Images of data collection of the crystal of the CMS-SH3A:YCbl-b$_{903-914}$ complex with a high resolution limit of 1.4Å (left) and 3.5Å (right).

In the 1.69-1.60Å shell, the Rsym increased from 44 to 70%, and the I/σ decreased from 1.8 to 1.1. By limiting the resolution to 1.65Å we obtained acceptable values of Rsym (56%) and I/σ (1.4) in the last shell, and still reached a 99.9% completeness.

Molecular replacement using the CMS-SH3A:Cbl-b$_{902-912}$ structure (PDB 2J6F) as a search model in Molrep gave an R-factor of 57% and a correlation coefficient of 56.2%. A rigid body refinement in Molrep improved R values to 47.6% for the R-factor and 69.8% for the correlation coefficient (with a contrast of 10.87 with the following best solution).

Refinement was carried out following the same refinement protocol that was used for the solution of the complex with the CIN85-SH3A:Cbl-b902-912 complex. Given that both crystals are isomorphous with only an extra 3% of the residues in the new crystal, free flags were maintained from the previous CMS-SH3A:Cbl-b structure. This was done so that the new model would not be flawed and, additionally, comparison of the cross-validation of each structure would be consistent.

The model converged after 110 cycles resulting in an R-factor of 22% and a free R-factor of 27% (see Table V.5).

Table V.5. Crystallographic data collection and analysis of CMS-SH3A:YCbl-$b_{903-914}$ structure.

CMS-SH3A:YCbl-$b_{903-914}$	
Space group	I422
Unit Cell Dimensions (Å)	a=b=66.471, c=68.058
	α=β=γ=90º
Data range (Å)	47.57-1.65
Observations (unique)	118888 (3783)
Completeness (%) (last shell)	99.9 (98.3)
R_{sym} [a] (last shell)	0.059 (0.056)
Reflections F>0 (cross validation)	8955 (519)
Non-hydrogen atoms (solvent molecules)	646 (519)
R_{cryst} [b] (R_{free} [c])	0.218 (0.270)
R.m.s. bond length (Å)	0.014
R.m.s. bond angles (º)	1.597

[a] R_{sym} is the unweighted R value on I between symmetry mates.
[b] $R_{cryst} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$
[c] R_{cryst} is the crossvalidation R factor for 5.5% of reflections against which the model was not refined.

Validation of the structure with the MolProbity tool showed no outliers for rotamers and neither in the Ramachandran plot (see Figure V.65). All other geometrical parameters showed acceptable values.

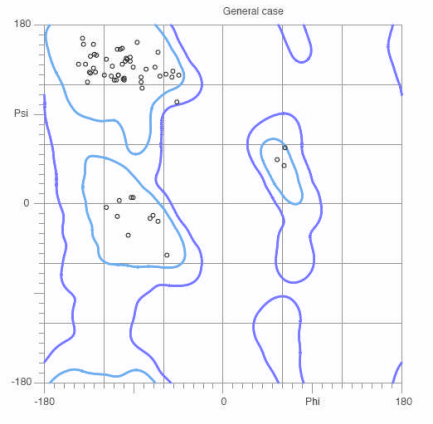


Figure V.65. Ramachandran plot of CMS-SH3A:YCbl-$b_{903-914}$ structure complex validation with the MolProbity web-based tool. All residues are in the allowed regions, with 97.2% of them in the favored regions.

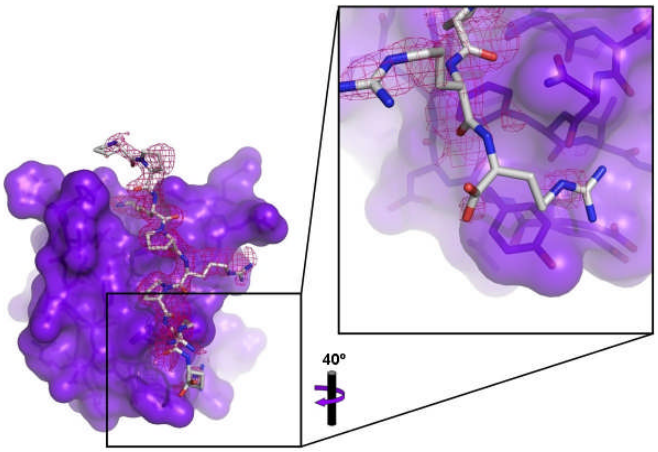


Figure V.66. CMS-SH3A:YCbl-b$_{903-914}$ structure. The SH3 molecule is shown as a purple surface and sticks. The Cbl-b peptide is represented in grey sticks in the class II-like orientatation only, with the electron density map around it contoured at 1.0σ.A zoomed view of Arginine 912 from the peptide is shown with a 40º rotation, where it can be readily observed that it is very disordered in the structure, and no continuity in the electron density is detected.

Not enough electron density was observed for the C-terminal residues of the peptide, disorder was detected starting in arginine 912 (see Figure V.66). Peptide was constructed for the same number of residues that for the previous structure, implying that interactions that take place in the first hydrophobic pocket of the class I-like recognition are not essential for the heterotrimer formation.

2.1.1.3. CMS-SH3A binding to different length Cbl-b derived peptides

Despite results obtained by the solution of the CMS-SH3A:Cbl-b heterotrimeric structures, ITC experiments were performed to compare binding affinity of CMS-SH3A to Cbl-b peptides of different lengths (YCbl-b$_{903-912}$ and YCbl-b$_{903-914}$, performed twice for each of them).

Our results showed a marginal increase in the affinity of 3.25-fold with the Cbl-b derived peptide two residues longer in its carboxy terminus (YCbl-b$_{903-914}$). Association constants (K_As) were calculated to be of $5.42x10^4$ and $1.76x10^5$ in a one-site binding system. These values correspond to dissociation constants (K_Ds) of 18.4μM and 5.7μM for the short (YCbl-b$_{903-912}$) and the long (YCbl-b$_{903-914}$) Cbl-b peptides, respectively. Both assays displayed a stoichiometry of 1:0.8 and the enthalpic contributions to ΔG were practically the same in both cases (ΔH was calculated to be -14.17 ± 1.23kcal/mol for the short peptide, and -12.08 ± 0.5kcal/mol for the long Cbl-b peptide).

These results show that the binding mechanism might not be affected by the length of the peptide, since the stoichiometry and enthalpic contribution are the same in both cases, and the affinity increase for the longer peptide is marginal the contribution of the extra C-terminal residues to the interaction is hence considered to be negligible.

DISCUSSION

"Science is the great antidote to the poison of enthusiasm and superstition."

Adam Smith

VI. Discussion

1. CIN85

A grown interest in the study of CIN85 has arisen with new discoveries on its implication in the regulation of important processes such as cell apoptosis, TKR downregulation or lysosomal degradation. As mentioned before, EGFR downregulation takes place upon CIN85-Cbl recognition.

We hoped that structural studies of CIN85:Cbl complexes would give us clues on the regulatory mechanism that governs the endocytosis of the EGFR. All CIN85 SH3 domains interact with an atypical motif in Cbl proteins and other cellular partners with different affinities (Kurakin, Wu et al.; Tibaldi and Reinherz 2003). This suggests intrinsic structural signatures in the SH3 responsible for the divergence in the affinities and selectivities.

1.1. CIN85 SH3 domains

Despite our efforts and other's on the crystallization of multiple CIN85 SH3 domain protein constructs in complex with Cbl derived peptides (Katrin Rittinger's personal communication) as well as multiprotein complexes with different CIN85 partners, the only suitable fragments for structural studies do not go further than the SH3 domains.

We have observed that crystallization of the individual SH3 domains of CIN85 depends directly on a low content of flexible fragments and high protein concentration. We proved that removal of large flexible fragments together with a drastic increment of the protein concentration (~10mM ≈ 90mg/ml) favored the ready crystallization of the SH3 domains, individually or in complex with the short polyproline peptides.

Sequence and structural alignment of the individual CIN85 SH3 domains shows a high degree of conservation in the residues that form the hydrophobic core and suggests a common fold. The structure of the individual CIN85 SH3 domains solved in the laboratory (SH3A and SH3C) show that all conserved core residues (leu18, Ile20 (usually Phe), Ile26, Ile29 (usually Val), Trp37, Gly39, Phe48 and Val53[8]), have the same side chain orientation in both SH3A and SH3C from CIN85. Residues that have been identified as highly conserved and critical for peptide binding in SH3 domains, such as Phe8, Tyr10, Trp36, Pro49, Asn51 and Phe52 are totally conserved in sequence and structural conformation in both Figure V.19 domains. Additional residues involved in peptide binding but with lower positional entropy adopt the same conformation in the hydrophobic pockets as well (Figure VI.1).

[8] Residue numbering corresponds to CIN85-SH3A.

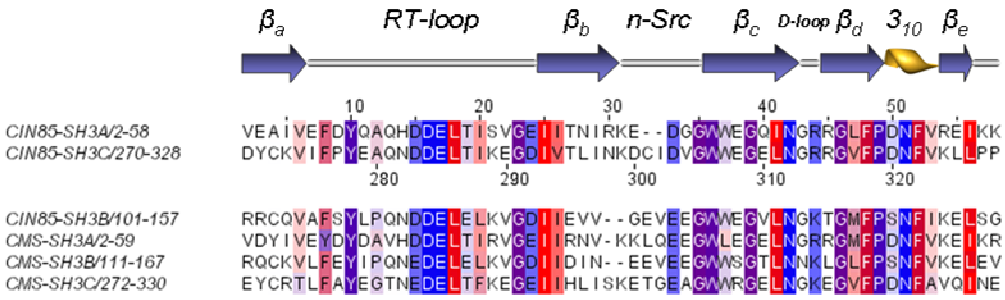


Figure VI.1. Sequence alignment of all CIN85 and CMS SH3 domains. Sequence alignment of CIN85 SH3A and SH3C are shown with residue numbers indicated. Secondary structural elements are illustrated on the top according to the CIN85 SH3 solved structures, where blue arrows represent β-strands, the yellow helix represents the 3_{10} helix, and grey lines represent the loops. CIN85-SH3B and CMS SH3 domains sequence alignment is shown at the bottom, colored by hydrophobic conservation.

In the overall structures of CIN85 SH3A and SH3C domains we observe a high conformational difference between the n-Src loops (Figure VI.2). A low sequence and structural conservation observed in the n-Src loop of SH3 domains in general correlates with a high variability in peptide conformation upon binding to the acidic pocket of SH3 domains, as reported by Larson and Davidson (Larson and Davidson 2000). This is explained by the fact that acidic residues of the RT and n-Src loops are responsible for binding specificity. Therefore, a higher variability among the SH3 sequence length and sequence conservation, as well as in structural arrangement, has been repeatedly reported.

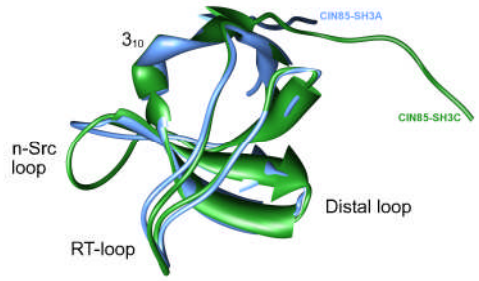


Figure VI.2. Structural alignment of CIN85-SH3A and CIN85-SH3C domains. CIN85 SH3A (blue) and SH3C (green) are superposed and represented in cartoons. The 3_{10} helix and the loops are indicated. Main structural differences are located in the n-Src loop.

The difference in the n-Src loop sequence and conformation between CIN85 SH3A and SH3C (see Figs. Figure VI.1 and Figure VI.2) might be the reason for the binding affinity divergence to the Cbl peptides observed in the fluorescence and ITC studies, along with IP and TAIS (Target-Assisted Iterative Screening) studies reported by Tibaldi and Reinherz (Tibaldi and Reinherz 2003) and Kurakin and co-workers (Kurakin, Wu et al. 2003), respectively.

1.2. CIN85-SH3A:Cbl-b structure

It has become evident that the interaction of c-Cbl or Cbl-b with adaptor proteins such as CIN85 and β-PIX is an important mechanism to modulate and fine-tune RTK endocytosis and degradation (Dikic and Giordano 2003; Lin, Yang et al. 2006; Schmidt, Husnjak et al. 2006). The same proline-arginine motif in the C-terminal portion of c-Cbl or Cbl-b is recognized by SH3 domains present in CIN85 and β-Pix proteins, suggesting that they may act in a competitive fashion with respect to Cbl. We reported in NSMB the crystal structure of the complex between a Cbl-b proline-arginine peptide and the N-terminal SH3 domain of CIN85 (CIN85-SH3A) described in the Results section V.1.1.3.1 (PDB ID 2BZ8), together with Daniela Jozic and Katrin Rittinger's structure of the complex between the Cbl-b peptide and β-PIX (Jozic, Cardenes et al. 2005) (PDB ID 2AK5). Notably, the Cbl-b peptide induces formation of a heterotrimeric complex consisting of two SH3 domains and one peptide molecule in both structures. Such heterotrimer formation appears to be mediated solely by the peptide binding, forming a structure arrangement that had not been described previously for SH3 domains.

Interactions made by the Cbl-b peptide with CIN85-SH3A and β-PIX are very similar and the peptide binds simultaneously resembling a class I and a class II orientations. Key to this unusual arrangement is the pseudo-symmetry of the peptide sequence. Such pseudo-symmetry allows two arginines, Arg904 and Arg911, to make equivalent hydrogen bonds with acidic residues in each SH3 domain from the heterotrimer, and it is also preserved in c-Cbl owing to the presence of the equivalent arginine residues, Arg822 and Arg829. Additionally, we observe that prolines 906 and 910 in the Cbl-b peptide make equivalent hydrophobic contacts with CIN85 phenylalanines 8 and 52 (β-Pix Phe15 and Tyr59) from each SH3 in the heterotrimer. Some unspecific interactions with carboxylic oxygens in the main chain are also found to be symmetrically related. We find hydrogen bonds established between the tryptophans 36 of each SH3 domain (β-Pix Trp43) and carboxylic oxygens of Pro905 and Arg909, which occupy the same position with respect to each SH3. Moreover, asparagines 51 from each CIN85-SH3A in the heterotrimer (β-Pix Asn58) hydrogen bond the Lys907 main chain oxygen, which occupies a central position in the pseudo-symmetric peptide.

Thanks to the analysis of the high resolution structure of the CIN85-SH3A:Cbl-b heterotrimer, the previously detected requirement of Arg911 for the CIN85:Cbl-b binding (Kowanetz, Szymkiewicz et al. 2003; Kurakin, Wu et al. 2003) can be rationalized by the extensive interactions between this residue and the SH3 domain. Such interactions include a series of hydrogen bonds between the carboxylic moieties of the acidic residues in the RT-loop of one of the SH3 domains ($Asp16_{SH3\ II}$ and $Glu17_{SH3\ II}$) and the guanidinium side chain of the C-terminal arginine, as well as the stacking of its aliphatic portion against Trp36 in CIN85. Arg904, by contrast, has not been previously recognized as important for complex formation with CIN85. Its presence seems to define the multimerization state of the SH3:peptide complexes and hence reveals why other ligands that also contain the PxxxPR motif, but lack the N-terminal arginine, for instance Pak2 or CD2, bind with a 1:1 stoichiometry (Jozic,

Cardenes et al. 2005). This was shown for CD2 in the CIN85-SH3A:$CD2_{324-333}$ structure (Results Section V.1.1.3.1.1). Furthermore, it suggests that other targets of CIN85 that lack the N-terminal arginine such as BLNK, the p85α subunit of PI3K, ZO-2, $TAF_{II}70$ or ASAP1 should preferentially bind in a 1:1 fashion (Figure VI.3).

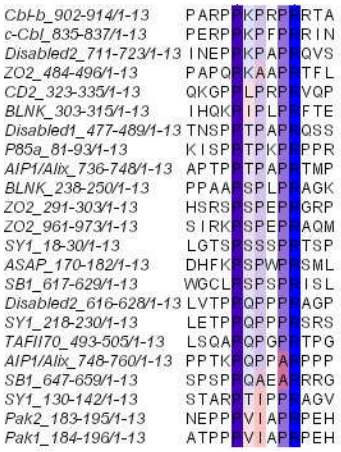


Figure VI.3. Alignment of interacting sequences from CIN85 natural partners. Several polyproline-rich sequences containing the PxxxPR motif from CIN85/CMS interacting molecules are aligned. Hydrophobic residues are shown in conservation gradient from *blue* to *red*. The prolines that interact with the hydrophobic pockets I in the CIN85-SH3A:Cbl-b heterotrimer (Cbl-b Pro906 and Pro910), and the C-terminal arginine responsible for the class II-like interaction with the RT-loop (Cbl-b Arg911) are totally conserved in all the partners. The N-terminal arginine responsible for the CIN85:Cbl-b heterotrimerization is only present in c-Cbl and Cbl-b proteins and two regions of the ZO-2 sequence.

Notably, ITC assays performed in collaboration with Katrin Rittinger showed that mutation of this Cbl-b N-terminal arginine (Arg904) does not appreciably change the apparent affinity of the peptide for CIN85-SH3A (2.5±1.0 and 4.4±1.1μM for the Cbl-b wild type and R904A mutant peptides, respectively), but changes the relative enthalpic and entropic contributions to ΔG (-14.25±0.5 and -7.9±1.6kcal/mol) (Jozic, Cardenes et al. 2005). In contrast, binding to β-PIX-$SH3_{C\text{-}term}$ is weakened ten-fold upon mutation of $Arg904_{Cbl\text{-}b}$ (from 14.0±3.5 to 140.0±30.0μM), indicating that, despite the similarity between the two structures, there must be differences in the mechanism of complex formation.

Our results suggest that Cbl proteins may additionally contribute to the formation of complexes needed for efficient EGF-receptor downregulation (Bartkiewicz, Houghton et al. 1999) by simultaneously interacting with two CIN85 SH3 domains.

Experiments in collaboration with Yonathan L. Deribe and Ivan Dikic showed that, *in vivo*, neither mutation of Cbl-b Arg904 nor that of Cbl-b Arg911 (or c-Cbl equivalent residues Arg822 and Arg829) abolished heterotrimerization (see Fig. 4 in Jozic *et al.* (Jozic, Cardenes et al. 2005); Appendix III, page 145 in this thesis). This was done by immunoprecipitation (IP) of a Flag-tagged CIN85 construct containing the three SH3 domains (FLAG-CIN85-3SH3), co-transfected together with wild type (WT) CIN85 and different c-Cbl or Cbl-b mutants, all of them impaired in their ability to homodimerize. Cbl-b-927 R904A mutant was impaired in its ability to mediate formation of a complex between full-length CIN85 and Flag-CIN85-3SH3.

Moreover, N- and C-terminal arginines in the c-Cbl/Cbl-b CIN85 binding motifs were mutated separately to determine their functionality in the binding to CIN85 and receptor downregulation. Additionally, they mutated both arginines simultaneously.

The amount of CIN85 co-precipitated with c-Cbl and Cbl-b proteins was quantified and showed that mutation of the C-terminal arginines, Arg829 and Arg911, reduced the interaction with CIN85 by approximately 80 and 60%, respectively, whereas mutation of Arg822 or Arg904 reduced binding by ~25% (see Fig. 4 in Jozic *et al.* (Jozic, Cardenes et al. 2005); Appendix III, page 145 in this thesis).

We know that efficient receptor endocytosis requires the formation of large protein complexes surrounding activated receptors (Dikic 2003; Szymkiewicz, Shupliakov et al. 2004). In order to investigate the role of these Cbl arginines essential for the binding to CIN85, the downregulation rate of EGF receptors in cells transfected with different Cbl mutants was measured. Mutation of either of the two arginine residues resulted in a substantial decrease of the degradation of EGF receptors, indicating that both play important roles in pathways controlling EGF-receptor downregulation (see Fig. 4 in Jozic *et al.* (Jozic, Cardenes et al. 2005); Appendix III, page 145 in this thesis).

These findings, together with biochemical and structural data on heterotrimerization of CIN85 by c-Cbl and Cbl-b, pointed to an important and previously unaccounted role of Cbl-b Arg904 and c-Cbl Arg822 in Cbl-mediated clustering of CIN85 and controlling endocytic trafficking of EGF receptors.

It has already been shown that CIN85 contributes to receptor internalization by clustering Cbl proteins via its three SH3 domains (Kowanetz, Szymkiewicz et al. 2003). However, our findings revealed yet another mechanism by which Cbl proteins can contribute to the formation of signal competent complexes implicated in receptor endocytosis.

Biochemical studies in cells clearly indicate that the proline-arginine motif in Cbl can simultaneously accommodate two SH3 domains from separate CIN85 molecules and could thereby promote the formation of higher-order Cbl:CIN85 complexes. At present we cannot exclude that adjacent SH3 domains in CIN85 may act in a similar manner and cooperate to bind a single ligand. Nevertheless, the identification of Arg904 in Cbl-b as an additional binding surface for SH3 domains explains how Cbl-mediated multimerization of CIN85 proteins occurs *in vivo*.

The physiological importance of these Cbl functions is likely to be relevant in the assembly of multiprotein complexes during receptor downregulation (Bartkiewicz, Houghton et al. 1999; Watanabe, Take et al. 2000; Kowanetz, Szymkiewicz et al. 2003). Notably, the β-PIX-SH3 domain binds in the same manner to Cbl proteins than CIN85. Thus, the competition of β-PIX and CIN85 for binding to the same surface of Cbl may represent the basis for dissociation of Cbl away from the activated EGF receptor upon activation of Cdc42:PIX complexes (Wu, Tu et al. 2003).

1.3. CIN85-SH3A:CD2 structure

Comparison of interactions of the N-terminal SH3 domain from CIN85 with peptides derived from other natural partners was necessary to have a deeper

knowledge of the interaction mode of this family of proteins, and the clustering mechanism promoted by SH3:ligand complex formation.

Analysis of the CIN85-SH3A:Cbl-b structure predicted a 1:1 stoichiometry in the CIN85-SH3A:CD2 binding, given that the N-terminal arginine participating in the class I-like binding is missing in the CD2 motif. We crystallized the same CIN85 SH3 with a CD2 derived peptide ($CD2_{324-333}$) containing the proline-rich PxxxPR binding motif (K*G*PPLPRPRV, residues belonging to the motif are underlined) but lacking the N-terminal arginine (where there is a glycine instead, shown in italics in the peptide sequence).

The crystal structure showed a dimeric complex formed by an SH3 domain that recognized the CD2 peptide in a Class II-like orientation, as expected. The CD2 peptide adopts the same conformation that the Cbl-b peptide in the trimeric complex, except for the non-interacting N-terminal end, where it differs slightly. Interactions with the peptide are equivalent to those observed in the CIN85-SH3A:Cbl-b Class II-like interface. We observe that Arg332 in CD2 plays essentially the same role as Arg911 in Cbl-b, establishing hydrogen bonds with the conserved acidic residues Asp16 and Glu17 in the n-Src loop of the CIN85-SH3A.

This results further suggest that the N-terminal arginine found in the Cbl proteins and others such as the ZO-2 tumor suppressor, can play an important role in the heterotrimer formation. Hence, the multimeric complexes formed with these partners that do not contain the N-terminal arginine will be different than those in which the heterotrimer complex is formed. This differential complex formation might yield different or modulated biological responses of the proteins involved.

2. CMS

In order to understand the molecular details of the different members of the CIN85/CMS family of adaptor molecules we wanted to compare the interaction and crystal complexes of these two adaptor molecules with Cbl and CD2 peptides.

2.1. CMS-SH3A:Cbl-b forms a heterotrimeric complex with a differential recognition from CIN85-SH3A:Cbl-b

The CMS-SH3A:Cbl-b crystal structures were used for comparative analysis with the CIN85-SH3A structures. Superposition of CMS-SH3A:Cbl-b and CIN85-SH3A:Cbl-b structures revealed that the overall structure of CMS-SH3A:Cbl-b (PDB ID 2J6F) resembles the heterotrimer reported for CIN85-SH3A:Cbl-b (Jozic, Cardenes et al. 2005), where one SH3 domain binds the Cbl-b peptide in a class I-oriented fashion, whereas the second SH3 recognizes the peptide in the opposite orientation (class II). Such similarity in the overall structures of the ternary complexes provides a molecular explanation for a common role of Cbl in the molecular mechanism of CIN85/CMS-mediated down-regulation of RTKs. Cbl might

also participate in the clustering of CMS as reported for CIN85 and β-PIX (Jozic, Cardenes et al. 2005). The Cα of the SH3 individual domains overlapped well with one another (root mean square deviation, r.m.s.d., of 0.70Å for 57 residues). Main differences were located within the n-Src and Distal loops. Nevertheless, the overall CMS-SH3A:Cbl-b heterotrimer showed some important differences with respect to CIN85-SH3A:Cbl-b. In particular, the orientation of the SH3 domains in CMS-SH3A:Cbl-b differs from that of the SH3 orientations in CIN85-SH3A:Cbl-b with a 56.4º shift with respect to each other, revealing a distinctive recognition of the Cbl-b-derived peptide for CMS or CIN85 (Figure VI.4).

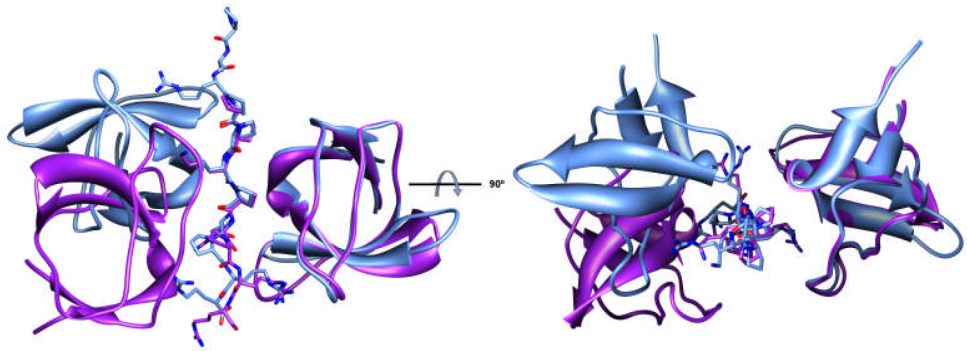


Figure VI.4. Comparison of the CIN85-SH3A:Cbl-b and CMS-SH3:Cbl-b ternary complexes. A ribbon representation of the SH3A domains of CIN85 (blue) and CMS (purple) in complex with the Cbl-b derived peptide (represented in sticks). The SH3A from CMS binding in a class II-like fashion was generated applying crystal symmetry operators. Complexes were superposed on SH3 II (class II-like; left). On the left panel, the complex is oriented with the peptides in a vertical position, with the N-terminus at the top. The complexes are rotated 90º towards the viewer to give the image on the right panel. Conformation of side chains of the peptide residues that bind in a class II-like orientation is essentially the same in both complexes.

The interactions corresponding to the class II-like orientation of the peptide (CMS-SH3A$_{SH3\ II}$:Cbl-b) are conserved with respect to CIN85 (CIN85-SH3A$_{SH3\ II}$:Cbl-b). As we have mentioned previously, Arg-911 determines the class II-like orientation, and its importance in (CMS/CIN85):Cbl interaction has already been described and proven (Kowanetz, Szymkiewicz et al. 2003; Kurakin, Wu et al. 2003). The conserved heterotrimeric arrangement suggests it also plays a critical role in CMS-SH3A:Cbl-b class II-like interaction. The only difference in the class II-like recognition between both structures concerns Cbl-b Arg-909, which hydrogen bonds with Glu35 from CMS and Asp33 in CIN85. However, significant differences in the interaction surface between Cbl-b and CMS-SH3A or CIN85-SH3A are observed in the class I-like orientation.

Residues forming the two hydrophobic binding pockets are highly conserved between CMS-SH3A and CIN85-SH3A, yet the peptide residues interacting with their SH3 I domains (class I-like) are different. Most differences are found in the acidic pocket formed by the n-Src and RT loops. Comparison of the CMS-SH3A (PDB code 2J6K, crystal structure solved in the laboratory by Mercedes Spínola-Amilibia) and CIN85-SH3A uncomplexed crystal structures show no conformational

changes in the n-Src and RT loops upon peptide binding (r.m.s.d. 0.3 and 0.7Å, respectively, according to the DaliLite web-based tool (Holm and Park 2000)). Therefore, differences observed between CMS-SH3A:Cbl-b and CIN85-SH3A:Cbl-b are not due to conformational changes induced upon peptide binding.

When compared to the CIN85-SH3A:Cbl-b structure, we observe that the N-terminus of the Cbl-b peptide adopts a different conformation that mostly reflects changes in the class I binding orientation. A significant three-residue shift of the Cbl-b peptide is observed (see Figure VI.5 and sections V.1.1.3.1 and V.2.1.1.1 for details).

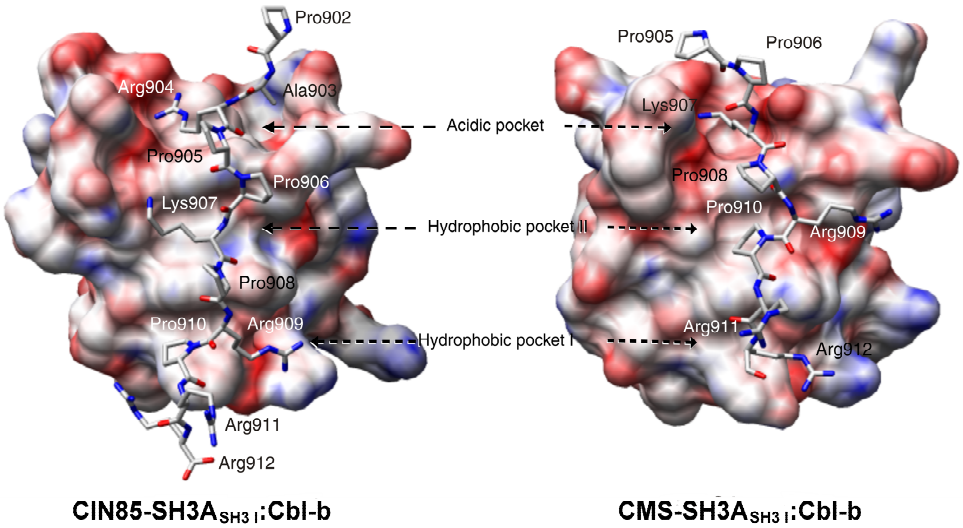


Figure VI.5. Comparison of the SH3 I:Cbl-b (class I-like) interface from CIN85-SH3A and CMS-SH3A. Electrostatic potential surfaces of SH3 domains from CIN85-SH3A:Cbl-b (*left*) and CMS-SH3A:Cbl-b (*right*) complexes that bind the peptide in a class I-like orientation are shown. The Cbl-b peptides are represented in *sticks*, and their residues are indicated. Binding pockets are indicated with *arrows*.

In CIN85, Asp16 and Glu17 from the RT loop hydrogen bond with Cbl-b Arg904, (critical in the formation of the CIN85-SH3A:Cbl-b ternary complex); however, in CMS-SH3A:Cbl-b, Arg904 does not participate in the interaction, and it is Lys907 that interacts with the RT loop in the acidic pocket (Figure VI.5).

In addition, Cbl-b Arg909 in CIN85-SH3A:Cbl-b occupies a similar position to Cbl-b Arg912 in the CMS-SH3A:Cbl-b structure, and hydrophobic pocket I is empty in CMS-SH3A$_{SH3\ I}$, because the synthesized peptide ended at Arg912. Therefore, we tested the affinity of a longer peptide (YCbl-b$_{903-914}$) by microcalorimetry, and we observed a marginal increase in the binding affinity. Even though there is certainly room for accommodation of the next residue in the Cbl-b sequence (Thr913), no electron density corresponding to Thr913 was observed in a 1.2Å refined CMS-SH3A:Cbl-b complex structure with YCbl-b$_{903-914}$ used as ligand. This might indicate that, although these extra residues might favor the binding and gently increase the affinity, the interaction with the SH3 domain is not strong and stable enough to be

observed in the crystal. It also suggests that hydrophobic pocket I does not play an important role in SH3A CMS mediated heterotrimer formation.

This three-residue shift in the class I orientation can be rationalized by sequence and structural differences within the respective n-Src loops, which have been reported to account for binding selectivity (Mayer 2001).With respect to the CIN85 sequence, a leucine is inserted in CMS at position 32 that significantly changes the overall conformation of the n-Src loop. In addition, Glu35 (Gly34 in CIN85) hydrogen bonds Arg909 from the peptide. The N-terminal end of the peptide faces opposite directions in each structure, and such displacement might be due to the change in the conformation of the n-Src loop, given that Glu34 in CMS does not allow the peptide to adopt a conformation similar to the one displayed in the CIN85-SH3A:Cbl-b complex.

From the analysis of the CMS-SH3A:Cbl-b structures we were again able to identify the importance of Cbl-b Lys907 and Arg911 for the interaction between the SH3 domain and its ligand. Biochemical experiments, similar to those performed for the CIN85:Cbl-b study, were carried out by Yonathan L. Derive in I. Dikic's group to see the importance of these residues in the context of full-length proteins expressed in cells and the formation of the heterotrimeric complex *in vivo*.

To this end, HA-tagged Cbl-b wt, Cbl-b K907A, Cbl-b R911A, and Cbl-b K907A/R911A point mutant proteins were overexpressed in HEK293T cells. We observed that both Cbl-b Lys907 and Arg911 play important roles in the interaction with CMS-SH3A, although the Cbl-b K907A mutation abrogated binding to a much lesser extent than Arg911. Double mutation of both Cbl-b Lys907 and Arg911 abolished the binding almost completely (see Fig. 2 in Moncalián *et al.* (Moncalian, Cardenes et al. 2006); Appendix III, page 154 in this thesis). Interaction through equivalent residues in c-Cbl was also checked with equivalent mutant proteins (c-Cbl K825A and c-Cbl R829A). Interestingly, a very similar profile of binding to Cbl-b was observed (see Fig. 2 in Moncalián *et al.* (Moncalian, Cardenes et al. 2006); Appendix III, page 154 in this thesis). The same effect was observed in the context of the full length proteins through precipitation of Cbl-b and c-Cbl wt and mutant proteins and detection of CMS (see Fig. 2 in Moncalián *et al.* (Moncalian, Cardenes et al. 2006); Appendix III, page 154 in this thesis). These experiments reproducibly showed the vital importance of these residues in binding to CMS.

The crystal structure of CMS-SH3A:Cbl-b suggested a different interaction pattern in class I recognition compared with CIN85-SH3A:Cbl-b. Cbl-b Lys907 in CMS-SH3A:Cbl-b occupies an equivalent position to Arg904 in the CIN85-SH3A:Cbl-b structure. Therefore, we wanted to compare the relative importance of Cbl-b Arg904 and Cbl-b Lys907 residues in binding to N-terminal SH3 domains from CMS and CIN85. Immunoprecipitation assays showed that GST/CMS-SH3A was able to bring down Cbl-b R904A as potently as wild type but pulled down Cbl-b K907A only weakly. In sharp contrast, GST/CIN85-SH3A was able to pull down both wild type Cbl-b and Cbl-b K907A mutant equally potently, with a slight reduction with respect to Cbl-b R904A in accordance with a previous observation (Jozic, Cardenes et al. 2005) (see Fig. 2 in Moncalián *et al.* (Moncalian, Cardenes et al. 2006); Appendix III, page 154 in this thesis).

The structure of CMS-SH3A:Cbl-b showed the formation of heterotrimers involving two CMS N-terminal SH3 domains and one Cbl peptide *in vitro*. To check whether the formation of CMS-SH3A:Cbl heterotrimers could also take place *in vivo*, co-immunoprecipitation studies similar to those carried out for CIN85 were performed, in which HEK293T cells were transfected with a FLAG-tagged construct of CMS containing only the three SH3 domain regions, Myc-tagged full-length CMS, and either c-Cbl-849 or Cbl-b-927 constructs that lack the ability to homodimerize. Cells overexpressing c-Cbl-849 or Cbl-b-927 showed the ability to co-precipitate full-length CMS using a FLAG antibody, indicating that Cbl is mediating the formation of the heterotrimer in cells, as it was shown for CIN85. Point mutants of c-Cbl or Cbl-b were otherwise significantly impaired in binding to CMS, corroborating the importance of Lys907 and Arg911 for the CMS:Cbl-b interaction that was inferred from the trimeric structure.

The importance of Cbl-b Lys907 is revealed by the fact that it is not conserved for other CMS ligands. CD2 displays a leucine (Leu328) at the equivalent position, which suggests that at least the class I type of interaction could be compromised upon CD2 binding to CMS-SH3A. In fact, crystals of a dimeric form of a 1:1 CMS-SH3A:CD2 complex on a class II type of interaction were obtained in the laboratory by Gabriel Moncalián. Accordingly, ITC studies of the binding of CMS-SH3A and a CD2 derived peptide showed a stoichiometry of 1:0.9 and a K_D constant of 26.6μM (data not shown, see Supplemental Materials). Unexpectedly, an additional heterotrimeric crystal was also obtained in an independent crystal form. The existence of the two molecular arrangements suggests a minimal energy difference between the heterotrimeric and dimeric states that might have some biological consequences in contrast to the predominant heterotrimeric complex formed by CMS-SH3A:Cbl-b (PDB ID 2J6F).

3. CIN85 and CMS SH3 binding consensus motifs

Biochemical studies performed in collaboration with Ivan Dikic's group provided *in vivo* confirmations for the observed structures and the residues involved in the binding between the SH3 domains and their ligands. Furthermore, we have shown a differential binding for SH3 domains of CMS and CIN85 to Cbl-b. This is in such a way that Cbl-b Arg904 is more important for binding and heterotrimer formation with CIN85-SH3A than for CMS-SH3A, whereas the opposite holds true for Cbl-b Lys907. In addition, it was also shown that Cbl mediates the formation of heterotrimers in the context of full-length CMS. Cbl-mediated multimerization of CMS and CIN85 proteins occurring *in vivo* suggests similar roles between both of them, contributing to the formation of large protein complexes required for receptor down-regulation. In spite of this, the differential Cbl molecular recognition between the two adaptor molecules brings about severe consequences in the arrangement of the SH3 domains. This distinct molecular arrangement might determine the overall clustering of the various proteins implicated in the endocytosis of the receptor with different implications in the biological roles.

The CMS and CIN85 SH3A complexed structures unambiguously indicate that residues involved in class I-like recognition are different from those involved in class II-like. Taken together, these results indicate that consensus sequences should indeed be considered independent for class I-like and class II-like recognition types. Prolines not involved in binding of one of the orientations may though help in providing the polyproline type II helix character by interaction of their δ carbons with the preceding residue and hence hindering that residue from adopting helical ψ angles (Bochicchio and Tamburro 2002). Nevertheless, the overall motif is similar to the one described by Kowanetz *et al.* (Kowanetz, Szymkiewicz et al. 2003) (PxP/AxPR), and another structure of CMS-SH3A:Cbl-b P908A at 1.9Å resolution solved by Gabriel Moncalián (unpublished results) confirmed that the same heterotrimerization is possible with Pro908 replaced by an alanine. It has been shown that some other hydrophobic residues could be present at this position without significantly affecting the interaction, such as isoleucine, found in the Pak polyproline motif (Hoelz, Janz et al. 2006). Thus, a more specific consensus motif for the heterotrimer formation could be narrowed down to RxPKxRPR in the case of CIN85 (where the lysine and the first arginine could most probably be substituted by a proline or a polar residue), and to PKxRPR for CMS (here the lysine and the second arginine are probably interchangeable) (see Figure VI.6).

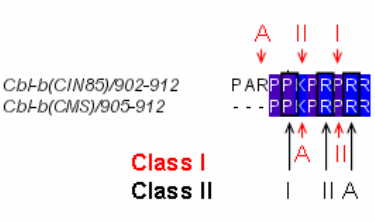


Figure VI.6. Cbl-b residues implicated in Class I-like and Class II-like recognition by SH3A from CIN85 and CMS. Cbl-b peptide residues observed in the CIN85-SH3A:Cbl-b and CMS-SH3A:Cbl-b ternary structures are sequence aligned and colored by hydrophobicity conservation, in a purple to blue gradient. Residues implicated in class II-like binding are indicated with black boxes. The positions that they occupy in the SH3 (Hydrophobic pocket I, (I), Hydrophobic pocket II (II) and Acidic pocket (A)) are indicated with arrows, black for class II-like and red for class I-like.

Analysis of the CIN85 and CMS binding motifs and the interactions that take place in each binding surface suggests that class II-like might be more stable than class I-like binding. Firstly, class II-like binding is conserved among most of the CIN85 and CMS natural partners (see Figure VI.3), and the interactions that are responsible for binding in this orientation are totally conserved in all of the structures presented in this study. Secondly, the number of interactions regarding class I-like recognition is lower than those of class II-like, as a matter of fact, hydrophobic pocket I is empty in the SH3 I from CMS-SH3A.

Additionally, we can predict from the peptide sequence if the formation of the ternary complex will take place. It has been demonstrated by mutagenesis studies that the presence of basic residues in the motifs determines the formation of the heterotrimer. As seen for the CIN85-SH3A:CD2 structure, the absence of the N-terminal arginine in the motif of a specific partner predicts a 1:1 binding in a class II-like orientation. Similarly, the absence of a basic residue in the position corresponding to Cbl-b Lys907 might hamper the class I-like recognition in CMS. However, the requirement of the basic residue for class I-like binding in CMS might be of less importance, as shown for CMS-SH3A:CD2 by G. M (PDB IDs 2J6O and

2J6I). This is also in agreement with the results obtained in the *in vivo* experiments, where the quantification of the contribution of each basic residue in the motif suggests a more predominant role of the arginine involved in the class II binding.

This effect has also been observed in binding to other partners that bear a +xPxxxP+ motif. Mutation of R309 to alanine in the **K**V**PT**V**P**P**R** motif from Sprouty2, which would correspond to the basic residue in the class II orientation, prevents interaction with CIN85 (Haglund, Schmidt et al. 2005).

4. Other examples of Ligand Induced SH3 Oligomerization (LISO)

At the time we solved the trimeric structures, the CIN85:Cbl and β-PIX:Cbl complexes were the only reported examples of ternary SH3-domain:peptide complexes in which the peptide induces trimerization and potentially oligomerization of its targets. Later on, we searched the database for possible examples of other SH3 ternary complexes to compare them with ours.

A few months after our report, another heterotrimeric SH3:polyproline complex in which two SH3 domains recognize the same polyproline peptide in opposite orientations was reported. An AMAP1 derived peptide ($AMAP1_{823-837}$; skKRPPPPPPGhkrt) was found to mediate heterotrimerization with two Cortactin SH3 domains (PDB ID 2DX1) (Hashimoto, Hirose et al. 2006).

Cortactin is a known CMS binding partner (Lynch, Winata et al. 2003), and this complex has been found to link the EGFR endocytosis and the Actin cytoskeleton. Moreover, AMAP1 mouse ortholog, ASAP1, was identified as a CIN85 effector interacting through an ASAP1 PxxxPR motif. Such interaction enhances the EGFR recycling and is believed to participate in the invasive machinery of several breast cancer cells (Kowanetz, Husnjak et al. 2004).

In contrast to the CIN85-SH3A:Cbl-b complex, the SH3 domains of Cortactin are related by a non-crystallographic twofold symmetry axis, positioning them antiparallely to each other but displaced 53.9º in respect to the CMS-SH3A:Cbl-b complex. The AMAP1 peptide adopts a left handed poplyproline type II (PPII) helix, in which a classical class I binding motif can be deduced (RxxPxxP). However, in the SH3 domain that recognizes the peptide in a class I orientation we find that both Arg826 and Lys825 form hydrogen bonds with the conserved acidic residues in the n-Src loop of Cortactin (Asp20 and Glu21), being the lysine the basic residue interacting more strongly instead of the predicted Arginine.

The authors found that, in the Cortactin-SH3:AMAP1 complex, two residues interact with both SH3 domains simultaneously. Pro829 is not only forming part of the Class I interface but also of the Class II-like contacts through interaction with Pro53 of both SH3 domains. Furthermore, Arg826, apart from the contact with the n-Src loop of the Class I oriented SH3, establishes an atypical interaction with the hydrophobic pocket I of the Class II-like recognizing SH3 by hydrogen bonding to

Tyr56. This peculiarity is what explains the anti-parallel orientation of the SH3 domains.

In the other hand, the Class II recognition of the AMAP1 peptide is not a typical one, since there is no basic residue in the C-terminus of the polyproline binding motif, unlike CIN85-SH3A Class II-like Cbl-b recognition.

As we have seen for CMS-SH3A:Cbl-b, the heterotrimeric crystal structure can be occluded by the crystallographic symmetry. It is possible that only through the generation of the crystallographically related molecules in the model, the biological heterotrimer is detected.

A fine search through the PDB database brings out another example of an SH3 heterotrimer that can only be generated by crystallographic symmetry, the Fyn-SH3:3BP-2 complex (PDB 1Fyn), in which the crystallographic axis is located perpendicular to the peptide.

The complex itself is atypical, since the peptide does not contain a classical RxxPxxP motif, instead, the peptide does not contain a basic residue to interact with the n-Src loop of the Fyn SH3. Unexpectedly, the residue interacting with the acidic pocket is a proline (3BP-2 Pro202).

The 3BP-2 peptide itself adopts a general PPII structure with a kink at Ala203 (Musacchio, Saraste et al. 1994) that allows Pro202 to fit in the acidic groove. Strikingly, if we compare the Fyn SH3 domain with the N-terminal SH3 domains from CIN85 and CMS, there are not many residue changes that would provide the n-Src loop from Fyn with a much higher hydrophobic character that would explain the contact with a proline (see Figure VI.8). Nonetheless, conserved glutamic and aspartic residues responsible for hydrogen bonding to the basic residue usually present in the polyproline motif are also present in Fyn-SH3.

In the case of the Class II-like recognition, the acidic pocket is not interacting with any residue from the peptide. Hydrophobic pockets I and II establish typical contacts with prolines 205 and 208 respectively.

In the case of the Myosin tail region-interacting protein MTI1 (BBC1) SH3 domain complexed structure with the Proline-rich protein Las17 derived peptide, a heterotrimeric structure is also observed. However, the orientation of the SH3 domains, and thus, the recognition of the peptide, differs highly from the CIN85 and CMS SH3 heterotrimeric structures.

The structure of the peptide does adopt a PPII helix conformation and class II recognition corresponds to the canonical, since Las17 R360 is hydrogen bonding Asp23 in the n-Src loop, and hydrophobic grooves I and II are occupied by prolines 355 and 358 respectively.

On the other hand, Class I binding can be detected by the recognition of the Las17 Arg350 by the Asp23 in SH3 I. In this case, it is virtually a classical class I binding, except that there is an extra alanine residue in the motif (RxxxPxxP) that allows arginine to bend towards the n-Src loop.

4.1. Comparison of all LISO arrangements

Comparison of all ternary complexes shows that none of them superpose perfectly (Figure VI.7). We wanted to identify sequence and structural features common among all of the trimeric structures.

Class II binding motifs are the most conserved among the peptide sequences (see bottom of Figure VI.8). Most important difference is that hydrophobic groove II in CIN85 and CMS holds an arginine (instead of the proline or alanine typically found in SH3:peptide binding) that is facing the n-Src loop and hydrogen bonds Glu35 in CMS. We still refer to it as a hydrophobic pocket because it is comprised of several conseved hydrophobic residues and for historical reasons. If we inspect CIN85 and CMS SH3A structures, we observe that this is possible thanks to the fact that there is a non-bulky residue at the top of the n-Src loop, CIN85 Gly35 and CMS Gly36. This glycine is conserved among all SH3 domains of the CIN85/CMS family and it is also present in the Cortactin SH3 sequence (Figure VI.1 and Figure VI.8). Additionally, there are no interactions in the acidic pocket of Fyn and Cortactin SH3 II (Class II oriented SH3) domains, given that their binding peptides do not bear a basic residue at their C-termini.

Most diversity is found in the class I binding. Here, all SH3 recognize an N-terminal basic residue (Figure VI.8) except for Fyn-SH3. The presence of a tyrosine residue in Fyn-SH3 n-Src loop (Tyr132) allows Pro202 from 3BP2 to establish hydrophobic contacts, atypical in the SH3:peptide recognition. The most sticking abnormality is the substitution of the proline residue by Cbl-b Lys907 in hydrophobic pocket II of CIN85-SH3A. We could not find any structural difference that would justify the preference of this amino acid at this position. The only difference in the sequence with respect to the other SH3 domains that could affect binding of this lysine residue is the presence of CIN85 Gln13 (Figure VI.1), which weakly hydrogen bonds lysine N_z (3.9Å). However, this does not seem to affect the specificity of the recognition of this residue, given that β-Pix-SH3 interacts with Cbl-b in the same manner and holds a threonine residue at this position.

The hydrophobic pocket I in the class I binding, is interacting with a proline in all cases, except for CMS, which is empty as discussed previously. Moreover, only two residues typically separate n-Src and hydrophobic groove II interacting residues in class I peptides. This is the case only for CIN85 and CMS SH3A. However, all other peptides in the LISO structures hold four and three residues between the positive residue (or 3BP2 Pro202) and the first consensus proline (see bottom of Figure VI.8). In the case of Fyn and BBC1, this is possible provided that the second residues in the motif sequence are very small (3BP2 Ala203 and Las17 Gly351) and allows bending of the N-termini towards the acidic pocket. As discussed earlier, AMAP1 Arg826 is also interacting with the acidic residues of the SH3 n-Src loop.

It appears to be evident that in order for the peptide-mediated heterotrimer to be formed, the peptide needs to embed motifs resembling a classical class I and a classical class II in its sequence. This is possible only if the peptide sequence is pseudosymmetric. We have seen though, that class II is more stable in some cases (CIN85 and CMS) and class I in others (Fyn, Cortactin and BBC1).

Inspection of the SH3 sequences does not reveal any feature that provides an explanation for the potentiality for the Ligand Induced SH3 Oligomerization, given that important residues involved in peptide binding are all conserved, such as Tyr10, n-Src acidic residues (Asp16 and Glu17 in CIN85), Trp36, Pro49[9] . In spite of this, we have observed that variation in the sequence might tolerate interaction with residues that differ from the consensus sequence motif, as discussed for Gly35 in CIN85 or Gly36 in CMS, and Tyr132 in Fyn.

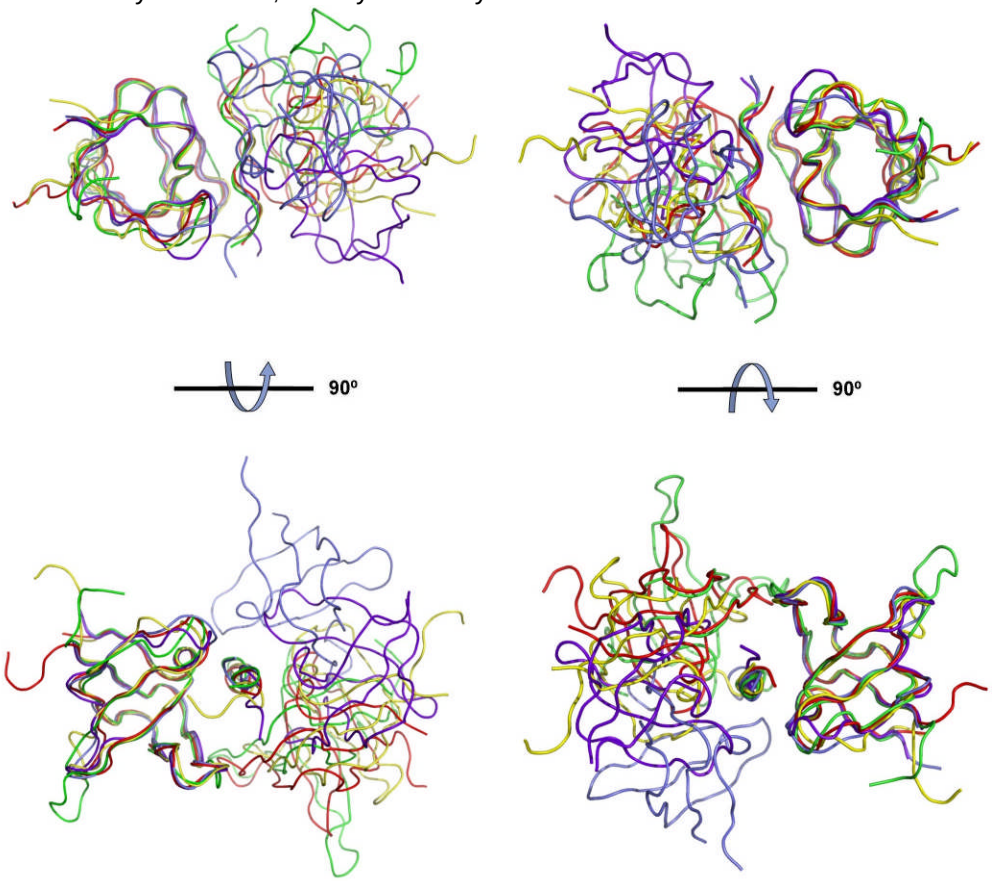
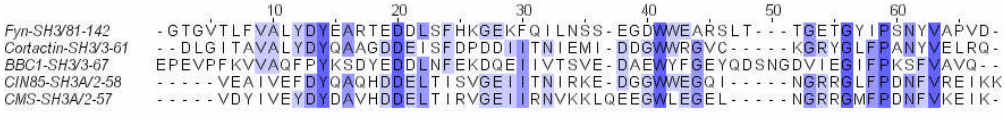


Figure VI.7. Comparison of SH3:polyproline heterotrimers. Ternary complexes of CIN85-SH3A:Cbl-b (blue), CMS-SH3A:Cbl-b (purple), Fyn-SH3:3BP-2 (yellow), Cortactin-SH3:AMAP1 (red) and BBC1:Las17 (green) are superposed on the SH3 binding the peptide in a class I-like orientation (left) or in a class II-like orientation (right). SH3 domains and polyproline peptides are represented as backbones. Two views of a 90º rotation are shown.



[9] Residues are numberer according to CIN85 sequence.

Figure VI.8. Sequence comparison of SH3:polyproline heterotrimer components. On the top, structural sequence alignment is shown for SH3 domains and for peptides in their class I and class II interfaces in the bottom. Residues are colored in a gradient of percent identity conservation. Basic amino acids interacting with the acidic pocket (A) are indicated with red boxes, and residues interacting with hydrophobic grooves I (I) and II (II) are indicated with black boxes. Black arrows indicate peptide residues at positions in the hydrophobic pockets not classical for class I and class II binding respectively.

In summary, we can predict the possibility of SH3 heterotrimerization from analysis of the peptide sequence, as we did for CIN85-SH3:CD2. Nevertheless, there are no clear SH3 sequence features that can help us predict the formation of such a molecular arrangement.

CONCLUSIONS

"Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world"

Louis Pasteur

VII.a Conclusions

- Taken together, we have identified a novel molecular recognition mechanism for SH3 domains where the same peptide is simultaneously recognized as Class I and Class II. Such molecular arrangement suggests that separate consensus motifs should be considered for each class orientation.

- This novel mechanism of Cbl mediated multimeric complex formation has provided Cbl with a new molecular functional role in the EGFR downregulation. Cbl acts as an adaptor that can accommodate multiple CIN85 molecules and consequently promote the formation of higher-order Cbl-CIN85 complexes.

- We have detected crucial residues in the proline-arginine motif of c-Cbl and Cbl-b involved in the formation of the ternary complex with the adaptor proteins CIN85 and CMS and the assembly of the endocytic machinery that governs downregulation of activated receptor tyrosine kinases.

- We have identified molecular features that might explain the slight differences in CIN85/CMS recognition of Cbl.

- Our analysis of the structures of other ligand induced SH3 heterotrimers has shown that formation of the trimeric complex in other cases is independent of the PxxxPR motif in the ligand.

VII.b Conclusiones

- Hemos identificado un novedoso mecanismo de reconocimiento molecular para los dominios SH3 donde el mismo péptide es reconocido simultáneamente como clase I y clase II. Dicha disposición molecular sugiere que se deben considerar diferentes motivos consenso de unión para cada orientación.

- Este novedoso mecanismo para la formación de complejos multiproteicos mediados por Cbl ha proporcionado una nueva función a Cbl en su papel en la regulación negativa del receptor EGFR. Cbl actúa entonces como una molécula adaptadora que puede acomodar varias moléculas de CIN85 y consequentemente promover la formación de complejos Cbl-CIN85 de mayor magnitud.

- Hemos detectado residuos cruciales en el motivo de poliprolina-arginina en c-Cbl y Cbl-b para la formación del complejo ternario con las proteínas adaptadoras CIN85 y CMS, y para el ensamblaje de la maquinaria endocítica que controla la regulación negativa de los receptores de tirosín-quinasa activados.

- Hemos identificado características moleculares que podrían explicar las diferencias diferencias en el reconocimiento de Cbl por CIN85 y CMS.

- Nuestro análisis de otras estructuras de heterotrímeros de SH3 inducidos por el ligando (*LISO: Ligand Induce SH3 Oligomerization*) ha mostrado que la formación de complejos ternarios en otros casos es independiente del motivo PxxxPR en el ligando.

BIBLIOGRAPHY

VIII. Bibliography

Aaronson, S. A. (1991). "Growth factors and cancer." Science **254**(5035): 1146-53.

Adzhubei, A. A. and M. J. Sternberg (1993). "Left-handed polyproline II helices commonly occur in globular proteins." J Mol Biol **229**(2): 472-93.

Bartkiewicz, M., A. Houghton, et al. (1999). "Leucine zipper-mediated homodimerization of the adaptor protein c-Cbl. A role in c-Cbl's tyrosine phosphorylation and its association with epidermal growth factor receptor." J Biol Chem **274**(43): 30887-95.

Ben-Bassat, A., K. Bauer, et al. (1987). "Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure." J Bacteriol **169**(2): 751-7.

Berry, D. M., P. Nash, et al. (2002). "A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling." Curr Biol **12**(15): 1336-41.

Blume-Jensen, P. and T. Hunter (2001). "Oncogenic kinase signalling." Nature **411**(6835): 355-65.

Bochicchio, B. and A. M. Tamburro (2002). "Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions." Chirality **14**(10): 782-92.

Bonifacino, J. S. and L. M. Traub (2003). "Signals for sorting of transmembrane proteins to endosomes and lysosomes." Annu Rev Biochem **72**: 395-447.

Borinstein, S. C., M. A. Hyatt, et al. (2000). "SETA is a multifunctional adapter protein with three SH3 domains that binds Grb2, Cbl, and the novel SB1 proteins." Cell Signal **12**(11-12): 769-79.

Breslauer, K. J., R. Frank, et al. (1986). "Predicting DNA duplex stability from the base sequence." Proc Natl Acad Sci U S A **83**(11): 3746-50.

Brett, T. J., L. M. Traub, et al. (2002). "Accessory protein recruitment motifs in clathrin-mediated endocytosis." Structure **10**(6): 797-809.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Cesareni, G., S. Panni, et al. (2002). "Can we infer peptide recognition specificity mediated by SH3 domains?" FEBS Lett **513**(1): 38-44.

Chen, B., S. C. Borinstein, et al. (2000). "The glioma-associated protein SETA interacts with AIP1/Alix and ALG-2 and modulates apoptosis in astrocytes." J Biol Chem **275**(25): 19275-81.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Crosetto, N., R. Tikkanen, et al. (2005). "Oncogenic breakdowns in endocytic adaptor proteins." FEBS Lett **579**(15): 3231-8.

Csiszar, A. (2006). "Structural and functional diversity of adaptor proteins involved in tyrosine kinase signalling." Bioessays **28**(5): 465-79.

De Filippis, V., A. Draghi, et al. (2007). "o-Nitrotyrosine and p-iodophenylalanine as spectroscopic probes for structural characterization of SH3 complexes." Protein Sci **16**(7): 1257-65.

Di Nardo, A. A., S. M. Larson, et al. (2003). "The relationship between conservation, thermodynamic stability, and function in the SH3 domain hydrophobic core." J Mol Biol **333**(3): 641-55.

Di Stasio, E., P. Bizzarri, et al. (2004). "A fast and accurate procedure to collect and analyze unfolding fluorescence signal: the case of dystroglycan domains." Biophys Chem **107**(2): 197-211.

Dikic, I. (2002). "CIN85/CMS family of adaptor molecules." FEBS Lett **529**(1): 110-5.

Dikic, I. (2003). "Mechanisms controlling EGF receptor endocytosis and degradation." Biochem Soc Trans **31**(Pt 6): 1178-81.
Dikic, I. and S. Giordano (2003). "Negative receptor signalling." Curr Opin Cell Biol **15**(2): 128-35.
Dikic, I., I. Szymkiewicz, et al. (2003). "Cbl signaling networks in the regulation of cell function." Cell Mol Life Sci **60**(9): 1805-27.
Downward, J., Y. Yarden, et al. (1984). "Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences." Nature **307**(5951): 521-7.
Drenth, J. (1994). Principles of Protein X-Ray Crystallography. New York, Springer-Verlag New York Inc.
Dustin, M. L., M. W. Olszowy, et al. (1998). "A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts." Cell **94**(5): 667-77.
Edelhoch, H. (1967). "Spectroscopic determination of tryptophan and tyrosine in proteins." Biochemistry **6**(7): 1948-54.
Edwards, A. M., C. H. Arrowsmith, et al. (2000). "Protein production: feeding the crystallographers and NMR spectroscopists." Nat Struct Biol **7 Suppl**: 970-2.
Eftink, M. R. (1997). "Fluorescence methods for studying equilibrium macromolecule-ligand interactions." Methods Enzymol **278**: 221-57.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.
Feng, S., J. K. Chen, et al. (1994). "Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions." Science **266**(5188): 1241-7.
Ferraro, E., D. Peluso, et al. (2007). "SH3-Hunter: discovery of SH3 domain interaction sites in proteins." Nucleic Acids Res **35**(Web Server issue): W451-4.
Ferreon, J. C. and V. J. Hilser (2004). "Thermodynamics of binding to SH3 domains: the energetic impact of polyproline II (PII) helix formation." Biochemistry **43**(24): 7787-97.
Finkel, T. a. G. J. S., Ed. (2003). Signal Transduction and Human Disease. Hoboken, New Jersey, John Wiley & Sons, Inc.
Freifelder, D. (1982). Physical Biochemistry: Applications to Biochemistry and Molecular Biology. New York, W. H. Freeman.
Giaccone, G. and J. A. Rodriguez (2005). "EGFR inhibitors: what have we learned from the treatment of lung cancer?" Nat Clin Pract Oncol **2**(11): 554-61.
Gill, S. C. and P. H. von Hippel (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Anal Biochem **182**(2): 319-26.
Gmeiner, W. H. and D. A. Horita (2001). "Implications of SH3 domain structure and dynamics for protein regulation and drug design." Cell Biochem Biophys **35**(2): 127-40.
Goulding, C. W. and L. J. Perry (2003). "Protein production in Escherichia coli for structural studies by X-ray crystallography." J Struct Biol **142**(1): 133-43.
Haglund, K., P. P. Di Fiore, et al. (2003). "Distinct monoubiquitin signals in receptor endocytosis." Trends Biochem Sci **28**(11): 598-603.
Haglund, K., M. H. Schmidt, et al. (2005). "Sprouty2 acts at the Cbl/CIN85 interface to inhibit epidermal growth factor receptor downregulation." EMBO Rep **6**(7): 635-41.
Haglund, K., S. Sigismund, et al. (2003). "Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation." Nat Cell Biol **5**(5): 461-6.

Hashimoto, S., M. Hirose, et al. (2006). "Targeting AMAP1 and cortactin binding bearing an atypical src homology 3/proline interface for prevention of breast cancer invasion and metastasis." Proc Natl Acad Sci U S A **103**(18): 7036-41.

Hoelz, A., J. M. Janz, et al. (2006). "Crystal structure of the SH3 domain of betaPIX in complex with a high affinity peptide from PAK2." J Mol Biol **358**(2): 509-22.

Holm, L. and J. Park (2000). "DaliLite workbench for protein structure comparison." Bioinformatics **16**(6): 566-7.

Hutchings, N. J., N. Clarkson, et al. (2003). "Linking the T cell surface protein CD2 to the actin-capping protein CAPZ via CMS and CIN85." J Biol Chem **278**(25): 22396-403.

Jelesarov, I. and H. R. Bosshard (1999). "Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition." J Mol Recognit **12**(1): 3-18.

Jia, C. Y., J. Nie, et al. (2005). "Novel Src homology 3 domain-binding motifs identified from proteomic screen of a Pro-rich region." Mol Cell Proteomics **4**(8): 1155-66.

Jozic, D., N. Cardenes, et al. (2005). "Cbl promotes clustering of endocytic adaptor proteins." Nat Struct Mol Biol **12**(11): 972-9.

Kalinin, A., N. H. Thoma, et al. (2001). "Expression of mammalian geranylgeranyltransferase type-II in Escherichia coli and its application for in vitro prenylation of Rab proteins." Protein Expr Purif **22**(1): 84-91.

Kaneko, T., T. Kumasaka, et al. (2003). "Structural insight into modest binding of a non-PXXP ligand to the signal transducing adaptor molecule-2 Src homology 3 domain." J Biol Chem **278**(48): 48162-8.

Karamouzis, M. V., J. R. Grandis, et al. (2007). "Therapies directed against epidermal growth factor receptor in aerodigestive carcinomas." Jama **298**(1): 70-82.

Kirsch, K. H., M. M. Georgescu, et al. (1999). "CMS: an adapter molecule involved in cytoskeletal rearrangements." Proc Natl Acad Sci U S A **96**(11): 6211-6.

Kishan, K. V., M. E. Newcomer, et al. (2001). "Effect of pH and salt bridges on structural assembly: molecular structures of the monomer and intertwined dimer of the Eps8 SH3 domain." Protein Sci **10**(5): 1046-55.

Kowanetz, K., K. Husnjak, et al. (2004). "CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors." Mol Biol Cell **15**(7): 3155-66.

Kowanetz, K., I. Szymkiewicz, et al. (2003). "Identification of a novel proline-arginine motif involved in CIN85-dependent clustering of Cbl and down-regulation of epidermal growth factor receptors." J Biol Chem **278**(41): 39735-46.

Kriminski, S., M. Kazmierczak, et al. (2003). "Heat transfer from protein crystals: implications for flash-cooling and X-ray beam heating." Acta Crystallogr D Biol Crystallogr **59**(Pt 4): 697-708.

Kurakin, A. V., S. Wu, et al. (2003). "Atypical recognition consensus of CIN85/SETA/Ruk SH3 domains revealed by target-assisted iterative screening." J Biol Chem **278**(36): 34102-9.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-5.

Lakowicz, J. R. (1999). Principles of Fluorescence Spectroscopy. New York, Kulwer Academic/Plenum Publishers.

Langdon, W. Y. (1995). "The cbl oncogene: a novel substrate of protein tyrosine kinases." Aust N Z J Med **25**(6): 859-64.

Larson, S. M. and A. R. Davidson (2000). "The identification of conserved interactions within the SH3 domain by alignment of sequences and structures." Protein Sci **9**(11): 2170-80.

Le Roy, C. and J. L. Wrana (2005). "Clathrin- and non-clathrin-mediated endocytic regulation of cell signalling." Nat Rev Mol Cell Biol **6**(2): 112-26.

Leslie, A. G. (2006). "The integration of macromolecular diffraction data." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 48-57.

Lewitzky, M., M. Harkiolaki, et al. (2004). "Mona/Gads SH3C binding to hematopoietic progenitor kinase 1 (HPK1) combines an atypical SH3 binding motif, R/KXXK, with a classical PXXP motif embedded in a polyproline type II (PPII) helix." J Biol Chem **279**(27): 28724-32.

Li, S. S. (2005). "Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction." Biochem J **390**(Pt 3): 641-53.

Lim, W. A., F. M. Richards, et al. (1994). "Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains." Nature **372**(6504): 375-9.

Lin, Q., W. Yang, et al. (2006). "Measurement of epidermal growth factor receptor turnover and effects of Cdc42." Methods Enzymol **406**: 614-25.

Lovell, S. C., I. W. Davis, et al. (2003). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." Proteins **50**(3): 437-50.

Lupher, M. L., Jr., C. E. Andoniou, et al. (1998). "The c-Cbl oncoprotein." Int J Biochem Cell Biol **30**(4): 439-44.

Lynch, D. K., S. C. Winata, et al. (2003). "A Cortactin-CD2-associated protein (CD2AP) complex provides a novel link between epidermal growth factor receptor endocytosis and the actin cytoskeleton." J Biol Chem **278**(24): 21805-13.

Mayer, B. J. (2001). "SH3 domains: complexity in moderation." J Cell Sci **114**(Pt 7): 1253-63.

Mayer, B. J. and M. J. Eck (1995). "SH3 domains. Minding your p's and q's." Curr Biol **5**(4): 364-7.

McPherson, A. (1999). Crysllization of Biological Macromolecules. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

McRee, D. E. (1999). "XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density." J Struct Biol **125**(2-3): 156-65.

Moncalian, G., N. Cardenes, et al. (2006). "Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain." J Biol Chem **281**(50): 38845-53.

Mongiovi, A. M., P. R. Romano, et al. (1999). "A novel peptide-SH3 interaction." Embo J **18**(19): 5300-9.

Mosesson, Y., K. Shtiegman, et al. (2003). "Endocytosis of receptor tyrosine kinases is driven by monoubiquitylation, not polyubiquitylation." J Biol Chem **278**(24): 21323-6.

Musacchio, A., M. Saraste, et al. (1994). "High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides." Nat Struct Biol **1**(8): 546-51.

Narita, T., F. Amano, et al. (2001). "Assignment of SH3KBP1 to human chromosome band Xp22.1-->p21.3 by in situ hybridization." Cytogenet Cell Genet **93**(1-2): 133-4.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Crystallogr. A **50**: 157-163.

Nave, C. and E. F. Garman (2005). "Towards an understanding of radiation damage in cryocooled macromolecular crystals." J Synchrotron Radiat **12**(Pt 3): 257-60.

Nishikawa, R., X. D. Ji, et al. (1994). "A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity." Proc Natl Acad Sci U S A **91**(16): 7727-31.

Normanno, N., A. De Luca, et al. (2006). "Epidermal growth factor receptor (EGFR) signaling in cancer." Gene **366**(1): 2-16.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276:** 307-326.

Panni, S., L. Dente, et al. (2002). "In vitro evolution of recognition specificity mediated by SH3 domains reveals target recognition rules." J Biol Chem **277**(24): 21666-74.
Pawson, T. (2002). "Regulation and targets of receptor tyrosine kinases." Eur J Cancer **38 Suppl 5**: S3-10.
Peschard, P. and M. Park (2003). "Escape from Cbl-mediated downregulation: a recurrent theme for oncogenic deregulation of receptor tyrosine kinases." Cancer Cell **3**(6): 519-23.
Petrelli, A., G. F. Gilestro, et al. (2002). "The endophilin-CIN85-Cbl complex mediates ligand-dependent downregulation of c-Met." Nature **416**(6877): 187-90.
Ravelli, R. B. and E. F. Garman (2006). "Radiation damage in macromolecular cryocrystallography." Curr Opin Struct Biol **16**(5): 624-9.
Ryan, P. E., G. C. Davies, et al. (2006). "Regulating the regulator: negative regulation of Cbl ubiquitin ligases." Trends Biochem Sci **31**(2): 79-88.
Salomon, D. S., R. Brandt, et al. (1995). "Epidermal growth factor-related peptides and their receptors in human malignancies." Critical Reviews in Oncology/Hematology **19**(3): 183-232.
Sambrook, J. and D. W. Russell (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.
Sawasdikosol, S., J. C. Pratt, et al. (2000). "Adapting to multiple personalities: Cbl is also a RING finger ubiquitin ligase." Biochim Biophys Acta **1471**(1): M1-M12.
Schlessinger, J. (2002). "Ligand-induced, receptor-mediated dimerization and activation of EGF receptor." Cell **110**(6): 669-72.
Schmidt, M. H. and I. Dikic (2005). "The Cbl interactome and its functions." Nat Rev Mol Cell Biol **6**(12): 907-19.
Schmidt, M. H., K. Husnjak, et al. (2006). "Cbl escapes Cdc42-mediated inhibition by downregulation of the adaptor molecule betaPix." Oncogene **25**(21): 3071-8.
Seidman, C. E. (1997). UNIT 1.8 Introduction of Plasmid DNA into Cells, Basic Protocol 1. Current Protocols in Molecular Biology. F. M. A. R. B. R. E. K. D. D. M. J. G. S. J. A. S. K. Struh, John Wiley & Sons, Inc.**:** 1.8.1-1.8.10.
Sheldrick, G. M., T. R. Schneider, et al. (1997). [16] SHELXL: High-resolution refinement. Methods in Enzymology, Academic Press. **Volume 277:** 319-343.
Sherman, F., J. W. Stewart, et al. (1985). "Methionine or not methionine at the beginning of a protein." Bioessays **3**(1): 27-31.
Sigismund, S., T. Woelk, et al. (2005). "Clathrin-independent endocytosis of ubiquitinated cargos." Proc Natl Acad Sci U S A **102**(8): 2760-5.
Soubeyran, P., K. Kowanetz, et al. (2002). "Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors." Nature **416**(6877): 183-7.
Stevens, R. C. (2000). "High-throughput protein crystallization." Curr Opin Struct Biol **10**(5): 558-63.
Szymkiewicz, I., O. Shupliakov, et al. (2004). "Cargo- and compartment-selective endocytic scaffold proteins." Biochem J **383**(Pt 1): 1-11.
Take, H., S. Watanabe, et al. (2000). "Cloning and characterization of a novel adaptor protein, CIN85, that interacts with c-Cbl." Biochem Biophys Res Commun **268**(2): 321-8.
Thien, C. B. and W. Y. Langdon (2001). "Cbl: many adaptations to regulate protein tyrosine kinases." Nat Rev Mol Cell Biol **2**(4): 294-307.
Thomas, L. (1974). The Lives of a Cell: Notes of a Biology Watcher. New York, Viking.
Tibaldi, E. V. and E. L. Reinherz (2003). "CD2BP3, CIN85 and the structurally related adaptor protein CMS bind to the same CD2 cytoplasmic segment, but elicit divergent functional activities." Int Immunol **15**(3): 313-29.

Tickle, I. J., R. A. Laskowski, et al. (1998). "Rfree and the rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement." Acta Crystallogr D Biol Crystallogr **54**(Pt 4): 547-57.

Tong, A. H., B. Drees, et al. (2002). "A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules." Science **295**(5553): 321-4.

Tronrud, D. E. (2004). "Introduction to macromolecular refinement." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2156-68.

Vieira, A. V., C. Lamaze, et al. (1996). "Control of EGF receptor signaling by clathrin-mediated endocytosis." Science **274**(5295): 2086-9.

Viguera, A. R., J. L. Arrondo, et al. (1994). "Characterization of the interaction of natural proline-rich peptides with five different SH3 domains." Biochemistry **33**(36): 10925-33.

Viguera, A. R., J. C. Martinez, et al. (1994). "Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition." Biochemistry **33**(8): 2142-50.

Wallace, R. B., J. Shaffer, et al. (1979). "Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch." Nucleic Acids Res **6**(11): 3543-57.

Wang, P. L., B. K. Lo, et al. (2005). "Generating molecular diversity by homologous recombination in Escherichia coli." Protein Eng Des Sel **18**(8): 397-404.

Warburg, O. and W. Christian (1942). "Isolierung und Kristallisation des Gärungsferments Enolase (Isolation and crystallization of enolase)." Biochem. Z. **310**: 384-421.

Watanabe, S., H. Take, et al. (2000). "Characterization of the CIN85 adaptor protein and identification of components involved in CIN85 complexes." Biochem Biophys Res Commun **278**(1): 167-74.

Waterman, H. and Y. Yarden (2001). "Molecular mechanisms underlying endocytosis and sorting of ErbB receptor tyrosine kinases." FEBS Lett **490**(3): 142-52.

Wu, W. J., S. Tu, et al. (2003). "Activated Cdc42 sequesters c-Cbl and prevents EGF receptor degradation." Cell **114**(6): 715-25.

Yeates, T. O. (1997). "Detecting and overcoming crystal twinning." Methods Enzymol **276**: 344-58.

Yee, A., X. Chang, et al. (2002). "An NMR approach to structural proteomics." Proc Natl Acad Sci U S A **99**(4): 1825-30.

Zarrinpar, A., R. P. Bhattacharyya, et al. (2003). "The structure and function of proline recognition domains." Sci STKE **2003**(179): RE8.

Zarrinpar, A., S. H. Park, et al. (2003). "Optimization of specificity in a cellular protein interaction network by negative selection." Nature **426**(6967): 676-80.

Zwart, P. H., G. G. Langer, et al. (2004). "Modelling bound ligands in protein crystal structures." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2230-9.

APPENDIXES

IX. Appendix I: Supplemental Material

X. Appendix II: Abbreviations

2xTY	Two times Tryptone-Yeast broth
Amp	Ampicillin
Ap	Ampicillin
Cbl	Cas-Br-M (murine) ecotropic retroviral transforming sequence
Cbl-b	Cas-Br-M (murine) ecotropic retroviral transforming sequence b
CD	Circular dicroism
CD2AP	CD2-associated protein
cDNA	Coding desoxyribonucleic acid
CIN85	Cbl-interacting protein of 85 kDa
CMS	Cas ligand with multiple SH3 domains
DNA	Desoxyribonucleic acid
E. coli	*Escherichia coli*
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ESRF	European Synchrotron Radiation Facility
IPTIG	Isopropyl β-D-1-thiogalactoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
Kn	Kanamycin
LB	Lysogeny broth
MR	Molecular replacement
NMR	Nuclear magnetic resonance
o/n	Over night
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein Data Bank
ppt	Precipitate
rpm	Revolutions per minute
s/n	Supernatan
SDS	Sodium Dodecyl Sulfate
SH3	Src Homology 3
SH3KBP1	SH3-domain kinase binding protein 1
TB	Terrific broth
TCP	Total cell pellet
TKR	Tyrosine kinase receptor
UV	Ultriviolet light

w/v	Weight to volume
β-MeOH	β-Mercapto ethanol

XI. Appendix III: Articles

“Scientific papers are important because without them scientists cannot get money from the government or from universities.”

E. R. Schulman

ARTICLES

nature structural & molecular biology

Cbl promotes clustering of endocytic adaptor proteins

Daniela Jozic[1,5], Nayra Cárdenes[2,5], Yonathan Lissanu Deribe[3,5], Gabriel Moncalián[2], Daniela Hoeller[3], Yvonne Groemping[4], Ivan Dikic[3], Katrin Rittinger[1] & Jerónimo Bravo[2]

The ubiquitin ligases c-Cbl and Cbl-b play a crucial role in receptor downregulation by mediating multiple monoubiquitination of receptors and promoting their sorting for lysosomal degradation. Their function is modulated through interactions with regulatory proteins including CIN85 and PIX, which recognize a proline-arginine motif in Cbl and thus promote or inhibit receptor endocytosis. We report the structures of SH3 domains of CIN85 and β-PIX in complex with a proline-arginine peptide from Cbl-b. Both structures reveal a heterotrimeric complex containing two SH3 domains held together by a single peptide. Trimerization also occurs in solution and is facilitated by the pseudo-symmetrical peptide sequence. Moreover, ternary complexes of CIN85 and Cbl are formed *in vivo* and are important for the ability of Cbl to promote epidermal growth factor receptor (EGFR) downregulation. These results provide molecular explanations for a novel mechanism by which Cbl controls receptor downregulation.

Ligand binding to receptor tyrosine kinases (RTKs) causes receptor activation and initiates intracellular signaling cascades, which ultimately control biological responses such as cell proliferation, cellular adhesion or apoptosis[1]. To regulate the lifetime of this process, activated RTKs also trigger negative feedback loops that inhibit receptor activation or promote receptor endocytosis and lysosomal degradation[2]. The latter process is mediated by the formation of large protein complexes capable of removing activated receptors from the plasma membrane via clathrin-dependent and independent pathways[3–6]. In this process, c-Cbl and Cbl-b proteins play a crucial role by acting as ubiquitin ligases as well as endocytic adaptor molecules. Cbl-mediated multiple monoubiquitination of RTKs sorts the receptor-containing vesicle to a degradation route and thereby leads to termination of the RTK signal[7–9]. Upon cell stimulation, Cbl proteins get translocated to activated RTKs and act as multifunctional adaptor proteins, capable of recruiting additional endocytic regulatory proteins such as CIN85–endophilin complexes, which facilitate receptor endocytosis.

A major function of CIN85 arises from its ability to bind to Cbl proteins as well as other effectors involved in regulation of receptor endocytosis[10–12]. Complex formation between CIN85 and Cbl occurs via the selective binding of SH3 domains of CIN85 to an atypical proline-arginine motif (PXXXPR, where X is any amino acid) present in the C-terminal region of c-Cbl and Cbl-b, and in other CIN85 effectors[13,14]. Because CIN85 has three SH3 domains (SH3A, SH3B and SH3C; **Fig. 1**) that all recognize the same motif, it has been proposed that CIN85 acts as a platform for clustering of Cbl molecules, thereby contributing to efficient degradation of activated EGFRs[13]. The closest homolog of CIN85-SH3A (from now on referred to as CIN85A) outside of the CIN85/CMS protein family is β-PIX, which has 52% sequence identity (**Fig. 1**). Notably, β-PIX has previously been shown to bind to a region of p21-activated kinase (Pak) that also contains an atypical proline-arginine motif[15]. Functionally, PIX proteins are believed to act as guanine nucleotide exchange factors (GEFs) for the small GTPases Rac and Cdc42 (refs. 15,16), which regulate many cellular processes including cell polarity and motility, gene transcription and phagocytosis[17]. Recently, β-PIX/Cool-1 has been described as another Cbl-interacting protein that modulates EGFR downregulation[18]. It was shown that activated Cdc42–PIX complexes promote sequestration of Cbl from the EGFR and thereby block EGFR endocytosis[19]. The underlying mechanism has yet to be unraveled, but our recent findings imply that β-PIX binds to the proline-arginine motif in Cbl and thus may compete with CIN85 for binding to Cbl (data not shown).

SH3 domains are classic examples of protein interaction modules. They generally bind to proline-rich sequences containing a core PXXP motif, although other sequences may also be recognized[20–23]. Many ligands bind their respective SH3 domains in a polyproline II helix (PPII) conformation[24]. Owing to the two-fold rotational pseudo-symmetry of this conformation, peptides can bind a given SH3 domain in either one of two opposite directions governed by the location of a positively charged residue, usually arginine, which precedes or follows the core PXXP motif[25,26]. SH3 domains associate with their ligands with a 1:1 stoichiometry, and proteins containing multiple SH3 domains can either serve as adaptor proteins to colocalize multiple proteins or bind multivalent ligands and thereby

[1]Division of Protein Structure, National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK. [2]Signal Transduction Group, Structural Biology and Biocomputing Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, E-28029 Madrid, Spain. [3]Institute of Biochemistry II, Goethe University Medical School, Theodor-Stern-Kai 7, 60590 Frankfurt a/M, Germany. [4]Abteilung Biomolekulare Mechanismen, Max-Planck-Institut fuer Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. [5]These authors contributed equally to this work. Correspondence should be addressed to K.R. (katrin.rittinger@nimr.mrc.ac.uk), J.B. (jbravo@cnio.es) or I.D. (ivan.dikic@biochem2.de).

Received 3 June; accepted 13 September; published online 9 October 2005; doi:10.1038/nsmb1000

increase the affinity and specificity of a given interaction. An exception is the so-called 'SuperSH3' domain, first noted in the NADPH oxidase subunit p47^{phox}, in which two tandem SH3 domains cooperate to bind a single ligand[27]. This configuration seems to constitute a novel protein interaction module that provides greater flexibility for the recognition of multiple, diverse ligands.

To understand the functional role of complex formation between Cbl and CIN85 or β-PIX, we investigated these interactions *in vitro* and *in vivo*. Crystal structures of complexes between a Cbl-b peptide and SH3 domains of either human CIN85 or rat β-PIX reveal that, to our surprise, the Cbl peptide can interact simultaneously with two molecules of either SH3 domain. We demonstrate by isothermal titration calorimetry (ITC) that heterotrimerization also occurs in solution and is dependent on the presence of an N-terminal arginine in the Cbl-b peptide. Structural analysis of these complexes shows that they are distinct from those formed by SuperSH3 domains and explains why other targets of CIN85 or PIX do not support heterotrimerization. Finally we provide evidence that Cbl-induced binding to two CIN85 molecules also occurs *in vivo*, indicating a previously unknown role of Cbl in mediating formation of larger protein complexes involved in regulation of receptor endocytosis.

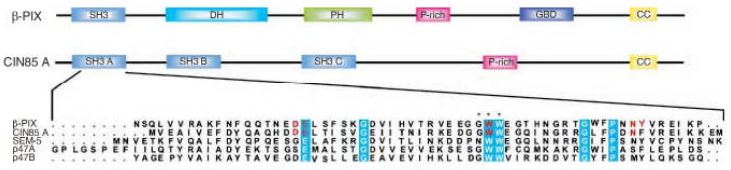


Figure 1 Domain structure of β-PIX and CIN85 and sequence alignment of SH3 domains. Schematic of the overall domain structure of β-PIX and CIN85 (DH, Dbl homology domain; PH, pleckstrin homology domain; P-rich, proline rich; GBD, Git-binding domain). CIN85 and β-PIX are both believed to multimerize via their respective coiled-coil (CC) domains (yellow). The sequence alignment was carried out among the SH3 domains of CIN85A, β-PIX, p47^{phox} and Sem-5. Red, residues that make contacts with the Cbl peptide; blue, completely conserved residues; asterisks, position of the Gly-Trp-Trp motif.

RESULTS

Overall structures of trimeric SH3–peptide complexes

Structures of complexes of CIN85A and the β-PIX SH3 domain each in complex with proline-arginine motif–containing peptides derived from Cbl-b were solved by molecular replacement and refined at resolutions of 2.0 and 1.85 Å, respectively. Each protein-peptide assembly crystallizes as a ternary complex that is composed of two SH3 domains and a single Cbl-b peptide (**Fig. 2a,b**). Both SH3 domains adopt the typical SH3 domain fold, a five stranded β-sandwich, and overlap with one another with r.m.s. deviations between Cα atoms of 0.8–1.2 Å. In each of the heterotrimeric complexes, two SH3 domains contact the Cbl-b peptide, which itself adopts a classical PPII helical conformation, through their conserved ligand-binding surfaces. However, one SH3 domain engages the peptide in a class I interaction (defined as SH3A), whereas the other binds on the opposite side of the peptide ligand in a class II orientation (defined as SH3B). The peptide is thus sandwiched by the

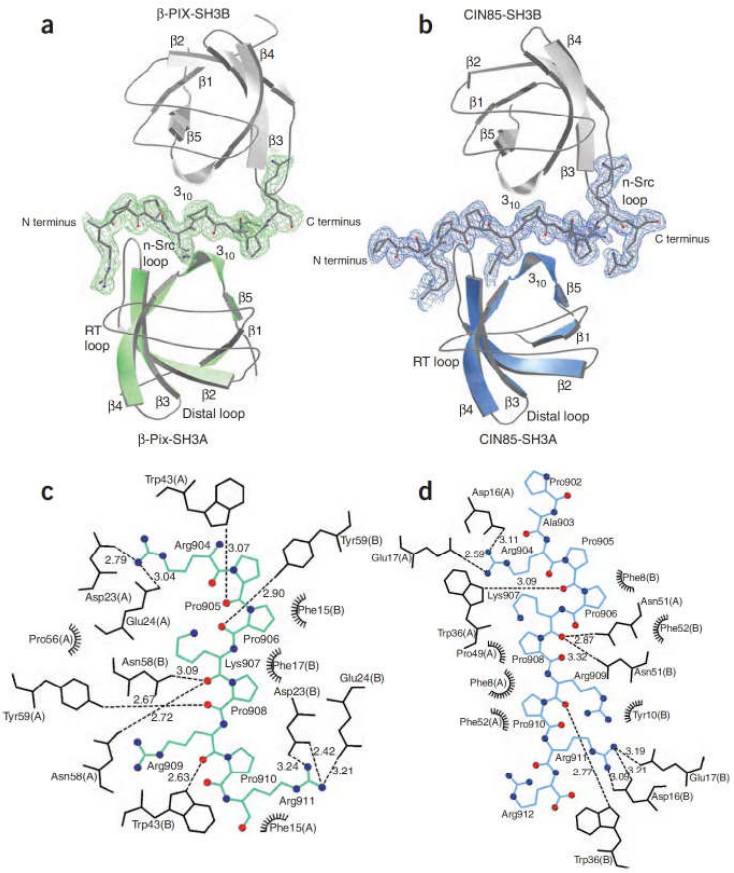


Figure 2 Overall structure of the β-PIX–Cbl-b and CIN85A–Cbl-b heterotrimeric complexes and details of the protein-peptide contacts. (**a**,**b**) The biologically relevant heterotrimeric complex of β-PIX and Cbl-b (**a**, green) and the trimeric complex of CIN85A and Cbl-b (**b**, blue). The electron density map around the peptide is contoured at 2.5 σ. The secondary structural elements and position of the RT, n-Src and distal loops are indicated. Both structures are shown in the same orientation to highlight their similarity. The peptide in the CIN85A complex shows additional density for residues Ala903, Pro902 and Arg912, which are disordered in the PIX complex. (**c**,**d**) A schematic representation of the protein-peptide contacts between the Cbl peptide and β-PIX (**c**) and CIN85A (**d**) . Green, PIX–Cbl-b peptide; blue, CIN85–Cbl-b peptide; black, residues from β-PIX and CIN85A; dashed lines, hydrogen bonds (labeled with peptide-protein distances in Å); black rays, hydrophobic interactions. SH3 domains are indicated in parentheses.

two SH3 domains in the ternary complex and there are no direct SH3-SH3 contacts in the CIN85A–Cbl-b structure. In the PIX–Cbl-b complex, direct contacts bury 73 Å^2 and are made through van der Waals contacts between PIX Asn58 from each domain.

Peptide binding and heterotrimerization

Complex formation between the Cbl-b peptide and either PIX SH3 domain buries 771 Å^2 (SH3A) or 776 Å^2 (SH3B). These values are comparable to those of the equivalent CIN85A–Cbl-b complex (928 and 715 Å^2, respectively) and generally to those of other SH3 domain–peptide structures. Many hydrogen bonds are formed by backbone carbonyl oxygens of the peptide to side chains of either SH3 domain (**Fig. 2c,d**). Specific interactions via side chains are only made by Arg904 and Arg911 from Cbl-b at the extreme N and C termini of the peptide with Asp23 and Glu24 of each SH3 domain of PIX or Asp16 and Glu17 of CIN85A, respectively. Hydrophobic contacts are made by Pro906 and Pro910 from Cbl-b with conserved residues from each SH3 domain in both structures, whereas Pro908 is surrounded by hydrophobic residues from SH3B of the PIX or CIN85A complexes.

Lys907 from the Cbl peptide plays a central role in both complexes. It is the only residue that makes simultaneous contacts with both SH3 domains through hydrogen bonds from its backbone carbonyl to the side chain amides of each, Asn58 in PIX–Cbl-b and Asn51 in CIN85A–Cbl-b. These contacts are pseudo-symmetrical and the backbone carbonyl of Lys907 constitutes the pseudo dyad axis of these interactions. Accordingly, equivalent interactions are established by Arg904 and Arg911 as well as Pro905 and Arg909 from Cbl-b (**Fig. 2**). The pseudo-symmetry of interactions is slightly disrupted around Pro906, which makes a hydrogen bond analogous to that of Pro908 (to Tyr59 from PIX), while at the same time being in a similar hydrophobic environment as Pro910. The main difference between the PIX–Cbl-b and CIN85A–Cbl-b complexes is that the residue corresponding to Tyr59 in PIX is a phenylalanine in CIN85A and hence there is no hydrogen bond formed with the backbone carbonyls of Pro906 and Pro908.

Heterotrimeric complex formation in solution

The heterotrimeric CIN85A–Cbl-b and PIX–Cbl-b complexes described here are, to our knowledge, the first examples of peptide-induced dimerization of SH3 domains. To ensure that this is a general process that occurs in solution and is not solely an artifact of crystallization, we analyzed the SH3 domain–peptide interactions by ITC. Binding of the Cbl-b peptide to PIX-SH3 resulted in a curve that could be fit to a single-site binding model with an apparent dissociation constant of 14 μM and a stoichiometry of 1:0.57 (**Table 1**), indicating that approximately only half the amount of peptide is required to form a complex with PIX. The same observation was made upon titration of CIN85A with the Cbl peptide: binding occurred with a stoichiometry of 1:0.57 and an apparent K_d of 2.5 μM (**Table 1** and **Fig. 3**). In contrast, titration of PIX-SH3 with a peptide derived from Pak1 (ref. 15) resulted in formation of complexes with a stoichiometry of 1:0.95 and a K_d of 3.2 μM (**Table 1**). Pak2 has been identified as a binding partner of CIN85 (ref. 14). However, the affinity of the Pak1-derived peptide, which is highly homologous to Pak2, for CIN85A was rather weak (K_d = 68 μM), supporting the notion that CIN85B may be the preferred ligand for Pak[14]. Nevertheless, given the high protein and peptide concentrations used in this titration, the data could be fit reliably and showed a 1:1 stoichiometry (**Table 1**, data not shown).

Table 1 Peptide binding of the SH3 domains of PIX and CIN85A

	n	K_d (μM)	ΔH (kcal mol^{-1})
β-PIX SH3 + Cbl-b	0.57 ± 0.1	14.0 ± 3.5	−8.35 ± 1.2
CIN85A + Cbl-b	0.57 ± 0.06	2.5 ± 1.0	−14.25 ± 0.5
β-PIX SH3 + Pak	0.95 ± 0.07	3.2 ± 0.3	−16.75 ± 0.3
CIN85A + Pak	0.92 ± 0.004	68.0 ± 2.0	−9.8 ± 0.1
β-PIX SH3 + mutant Cbl-b	0.75 ± 0.27	140.0 ± 30.0	−13.3 ± 5.7
CIN85A + mutant Cbl-b	1.24 ± 0.03	4.4 ± 1.1	−7.9 ± 1.6

All titrations apart from CIN85A with Pak and β-PIX SH3 with mutant Cbl-b were carried out between two and four times and the estimated error is the s.d. between measurements. The latter two titrations were carried out only once owing to the high protein and peptide concentrations required for these weak interactions. The error given for these titrations is the error of the respective fit.

Previous studies have indicated that Arg911 in Cbl-b (underlined in PXXXPR) is important for the interaction with CIN85 (refs. 13,14), an observation that can now be rationalized by the multiple hydrogen bonds formed between its guanidinium group and Asp16 and Glu17 from SH3B of CIN85A or Asp23 and Glu24 from SH3B of PIX, respectively. By contrast, Arg904, which makes equivalent contacts with SH3A from CIN85A and PIX, has not previously been implicated in complex formation. Our structures suggest that this residue may be crucial for peptide-induced dimerization, an idea supported by the observation that this arginine is absent in the Pak peptide, which binds either SH3 domain with a 1:1 stoichiometry. To test this hypothesis we measured binding of a mutant R904A Cbl peptide to CIN85A and PIX. Titration of CIN85A with this peptide clearly showed that stoichiometric amounts are required to saturate the SH3 domain with an affinity of 4.4 μM, further supporting our model (**Fig. 3** and **Table 1**). In contrast, mutation of Arg904 weakened the interaction with PIX to 140 μM, thereby making determination of the stoichiometry of the interaction somewhat unreliable. Nevertheless, considerably more than half an equivalent of mutant peptide is required to approach saturation, suggesting that this interaction also occurs with a 1:1 stoichiometry.

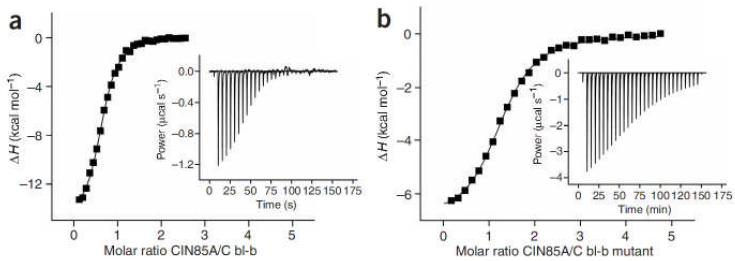


Figure 3 Isothermal titration calorimetry measurements of complex formation. (**a**,**b**) ITC of complex formation between CIN85A (30 μM) and the wild-type Cbl peptide (333 μM) (**a**) and between CIN85A (139 μM) and the mutant Cbl-b peptide (1,084 μM) (**b**). Insets show the raw data of each titration. The optimal fits for these two titrations give a K_d of 1.7 μM and an *n* of 0.61 for the wild-type peptide and a K_d of 3.6 μM and an *n* of 1.27 for the mutant R904A peptide. To better illustrate the different stoichiometries between these two titrations the *x*-axes are shown on the same scale. The dissociation constant determined for the CIN85A–Cbl-b complex is only an apparent K_d as it includes binding as well as trimerization.

The stoichiometry of the CIN85A–Cbl-b complex described in this manuscript differs from an earlier study, which reported a 1:1 complex[13]. At present we are not certain about the reasons for this discrepancy but believe that it is probably due to the experimental conditions and the size of the constructs used, which differ extensively between the two studies.

In vivo interactions between c-Cbl/Cbl-b and CIN85

Previous findings have indicated that the C-terminal leucine zipper (LZ) of c-Cbl mediates protein dimerization and thereby formation of larger complexes needed for efficient EGF-receptor downregulation[28]. The data presented here suggest that Cbl proteins may additionally contribute to the formation of such complexes by simultaneously interacting with two CIN85-SH3 domains. We first tested whether c-Cbl that also contains the pseudo-symmetrical proline-arginine motif can promote ternary complex formation with CIN85 *in vivo* by cotransfection of full-length CIN85 and Flag-CIN85-3SH3 with different c-Cbl mutants in HEK293T cells. Flag-CIN85-3SH3 (a construct containing the three SH3 domains of CIN85) is impaired in its ability to oligomerize because it lacks the coiled-coil motif and hence cannot form heterocomplexes with wild-type CIN85 (ref. 13). This makes it particularly useful in monitoring interactions between SH3 domains of CIN85 in the presence or absence of a potential linker such as c-Cbl. Accordingly, coprecipitation between CIN85 and Flag-CIN85-3SH3 was detected at very low levels in cells, probably owing to the presence of endogenous c-Cbl (**Fig. 4a**). By contrast, coprecipitation was increased upon coexpression with a c-Cbl deletion mutant lacking the LZ domain (c-Cbl-849), which retained its ability to bind to Flag-CIN85-3SH3 (**Fig. 4a**, top panel). This mutant was used to ensure that coprecipitation was a consequence of binding of two independent CIN85 molecules to a single c-Cbl molecule rather than of the interaction with homodimeric c-Cbl[28]. More importantly, a c-Cbl mutant (c-Cbl-827) lacking the LZ domain and the proline-arginine binding site did not interact with SH3 domains of CIN85 and thus did not promote coprecipitation between Flag-CIN85-3SH3 and wild-type CIN85 (**Fig. 4a**, bottom panel). Because the crystal structures and ITC studies presented here indicate that Arg904 in the Cbl-b peptide is crucial for allowing its heterotrimerization with CIN85-SH3A, we analyzed the role of this residue on the *in vivo* interaction between CIN85 and Cbl-b (**Fig. 4b**). A Cbl-b deletion mutant lacking the LZ/UBA domain (Cbl-b-927), which cannot homodimerize, interacted strongly with SH3 domains of CIN85, whereas Cbl-b-927 R904A was reduced, but not blocked, in its ability to coprecipitate with Flag-CIN85-3SH3 (**Fig. 4b**). Furthermore, the Cbl-b-927 R904A mutant was impaired in its ability to mediate formation of a complex between full-length CIN85 and Flag-CIN85-3SH3 (**Fig. 4b**). We noticed that c-Cbl was consistently more effective than Cbl-b in inducing trimerization with CIN85 (**Fig. 4a,b**). However, the

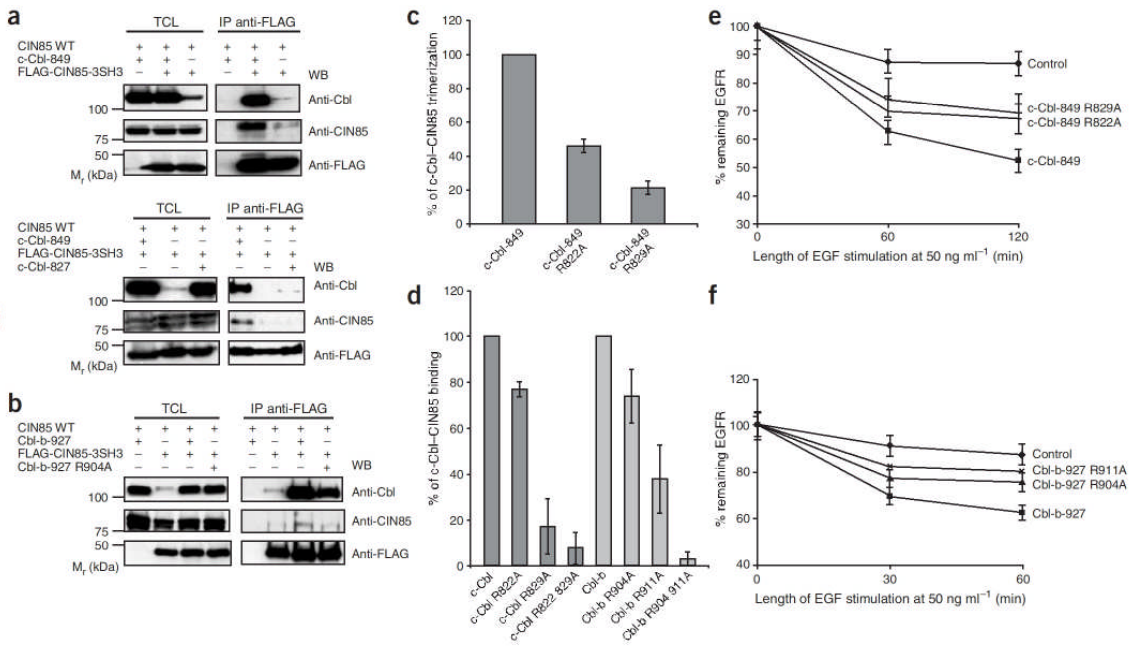


Figure 4 c-Cbl and Cbl-b mediate ternary complex formation with CIN85 *in vivo*. (**a**) HEK293T cells transfected with Flag-CIN85-3SH3, full-length CIN85 and c-Cbl-849 or c-Cbl-827 were lysed and immunoprecipitated by anti-Flag affinity agarose beads. Total cell lysates (TCL) and immunoprecipitates (IP) were subjected to immunoblotting (WB) with corresponding antibodies. (**b**) Cbl-b-927 or Cbl-b-927-R904A were transfected along with Flag-CIN85-3SH3 and full-length CIN85 in HEK293T cells and cell lysates were analyzed as described in **a**. (**c**) Transfection and immunoprecipitations were done as in **a** and **b**. Immunoblots from at least three independent experiments were then quantified to assess Cbl–CIN85 trimerization. (**d**) HEK293T cells were transfected with the indicated constructs and subjected to similar treatment as above. Coimmunoprecipitated c-Cbl or Cbl-b from at least three independent experiments was quantified by ImageJ[43]. (**e**) Mutation of Arg822 and Arg829 impairs the ability of c-Cbl-849 to promote EGFR degradation. Total EGFR levels were determined by flow cytometry in HEK293T cells expressing EGFR-GFP and the indicated c-Cbl mutants. (**f**) Mutation of Arg904 and Arg911 impairs the ability of Cbl-b-927 to promote EGFR degradation.

pseudo-symmetrical proline-arginine motif is almost identical in both Cbl isoforms, indicating that this effect must be caused by differences in other regions of the two proteins. Because of its stronger binding, we used c-Cbl to quantify the relative contributions of the two conserved arginines, Arg822 and Arg829, to interaction with and trimerization of CIN85. Mutation of either Arg822 or Arg829 considerably reduced their ability to promote simultaneous binding to full-length CIN85 and Flag-CIN85-3SH3 (**Fig. 4c**), supporting the conclusion that an intact pseudo-symmetrical proline-arginine motif is required to allow Cbl proteins to promote formation of the heterotrimeric complex *in vivo*. However, heterotrimerization was not completely abolished, suggesting that, in cells, either additional interacting partners of c-Cbl and CIN85 may indirectly contribute to formation of the heterologous complex or regions outside the core binding motif may further stabilize this complex.

To determine the functional role of the individual arginines in full-length c-Cbl and Cbl-b in binding to CIN85 and receptor downregulation, we separately mutated either Arg822 or Arg829 in c-Cbl, or Arg904 or Arg911 in Cbl-b; additionally, we mutated both of them simultaneously. The amount of CIN85 coprecipitated with c-Cbl and Cbl-b proteins was quantified and showed that mutation of Arg829 and Arg911 reduced the interaction with CIN85 by approximately 80 and 60%, respectively, whereas mutation of Arg822 or Arg904 reduced binding by ~25%. Mutation of both sites in either c-Cbl or Cbl-b largely abolished their coprecipitation with CIN85 (**Fig. 4d**). Last, we tested the functional importance of these arginines by measuring the downregulation rate of EGF receptors in cells transfected with different Cbl mutants. Expression of dimerization-deficient c-Cbl-849 or Cbl-b-927 in HEK293 cells enhanced EGF-induced degradation of EGF receptors (**Fig. 4e,f**). This effect was substantially reduced when either of the two arginine residues was mutated, indicating that both play important roles in pathways controlling EGF-receptor downregulation. These findings, together with our data on heterotrimerization of CIN85 by c-Cbl and Cbl-b (**Fig. 4a–c**), point to an important and previously unaccounted role of Cbl-b Arg904 and c-Cbl Arg822 in Cbl-mediated clustering of CIN85 and controlling endocytic trafficking of EGF receptors.

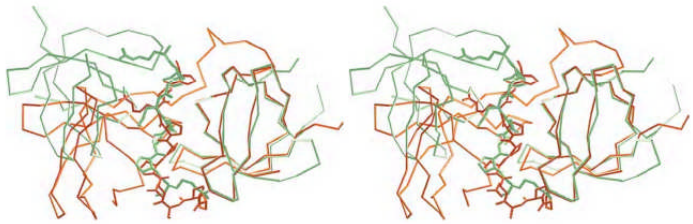


Figure 5 Comparison of the ternary β-PIX–Cbl-b complex with a SuperSH3 domain. Stereo representation of the overlaid backbones of SH3B from the β-PIX–Cbl-b complex (green) and the p47–p22^{phox} structure (orange).

DISCUSSION

It has become evident that the interaction of c-Cbl or Cbl-b with adaptor proteins such as CIN85 and PIX is an important mechanism to modulate and fine-tune RTK endocytosis and degradation[2]. The same proline-arginine motif in the C-terminal portion of c-Cbl or Cbl-b is recognized by SH3 domains present in both proteins, suggesting that they may act in a competitive fashion with respect to Cbl. We report here the crystal structures of complexes between a Cbl-b proline-arginine peptide and the SH3 domains of β-PIX and CIN85 (SH3A), respectively. Notably, the Cbl peptide induces formation of a heterotrimeric complex consisting of two SH3 domains and one peptide molecule in both structures. Interactions made by the Cbl peptide with β-PIX and CIN85A are very similar and the peptide binds simultaneously in class I and II orientations. Key to this unusual arrangement is the pseudo-symmetry of the peptide sequence, which allows two arginines, Arg904 and Arg911, to make equivalent hydrogen bonds with acidic residues in each SH3 domain. This pseudo-symmetry is also preserved in c-Cbl owing to the presence of Arg822 and Arg829. The previously shown requirement of Arg911 for complex formation can now be rationalized by the extensive interactions between this residue and the SH3 domain: the guanidinium side chain is involved in a network of hydrogen bonds with acidic residues in the RT loop while its aliphatic portion stacks against Trp43 in PIX and Trp36 in CIN85. Arg904, by contrast, has not been previously recognized as important for complex formation with CIN85. Its presence seems to define the multimerization state of the SH3–peptide complexes and hence indicates why other ligands that also contain the PXXXPR motif, but lack the N-terminal arginine, such as Pak, bind with a 1:1 stoichiometry. Furthermore, it suggests that other targets of CIN85 that lack the N-terminal arginine such as BLNK or CD2 should bind in a 1:1 fashion. Notably, mutation of this arginine does not appreciably change the apparent affinity of the peptide for CIN85A, but changes the relative enthalpic and entropic contributions to ΔG. In contrast, binding to PIX-SH3 is weakened ten-fold upon mutation of Arg904, indicating that despite the similarity between the two structures there must be differences in the mechanism of complex formation. Notably, *in vivo*, neither mutation of Arg822 nor that of Arg829 (the equivalent residues in c-Cbl) completely abolishes heterotrimerization. This may indicate that in the context of the full-length proteins, residues outside of the core motif might contribute to the stabilization of a heterotrimeric complex, or alternatively that additional interacting proteins may indirectly contribute to their coprecipitation by forming multimeric complexes.

Comparison with SuperSH3 domains

The PIX–Cbl and CIN85–Cbl structures are at present the only examples of ternary SH3 domain–peptide complexes in which the peptide induces dimerization and potentially oligomerization of its targets. The only other example of SH3–ligand complexes in which the ligand is shared between two SH3 domains is p47^{phox} in the autoinhibited and the p22^{phox}-bound state, a structural arrangement that has been termed the SuperSH3 domain[27]. However, in contrast to the structures described here, the SH3 domains of p47^{phox} are covalently linked and this link is required to allow the two SH3 domains to simultaneously bind the same peptide. Although the individual SH3 domains of the p47^{phox} and PIX or CIN85 structures overlap well with one another, an overlap of the trimeric PIX–Cbl structure with that of the dimeric p47–p22^{phox} complex shows that the second SH3 domain in each structure occupies a different position (**Fig. 5**). In the PIX–Cbl-b and CIN85A–Cbl-b complexes, the two SH3 domains come closest at Asn58 of PIX and Asn51 of CIN85, situated in a 3_{10}-helix in the C-terminal portion of the proteins (**Figs. 1** and **2**). In p47^{phox}, the

n-Src loops come closest, burying 579 Å^2 of solvent-accessible surface (through direct domain interactions) and contributing to the stability of the SuperSH3 domain conformation through hydrogen bonds across the interface. A GWW motif present in each n-Src loop acts as a signature motif of SuperSH3 domains[27]. The glycine is required for steric complementarity while the second tryptophan residue makes important interactions across the interface. Notably, PIX as well as all three SH3 domains of CIN85 contain this motif (**Fig. 1**). Nevertheless, their relative orientation in the heterotrimeric structures described here is different from that observed in p47^{phox} and residues of the n-Src loop do not contribute to trimerization. This suggests that the new structures presented here are not additional examples of SuperSH3 domains but rather constitute a novel mechanism by which SH3 domains can contribute to the formation of multiprotein complexes.

Cbl-mediated protein multimerization in endocytosis

Efficient receptor endocytosis requires the formation of large protein complexes surrounding activated receptors[4,29]. In many instances, specialized endocytic scaffold proteins engage multiple effectors and couple trafficking cargo with the clathrin coat, actin cytoskeleton and endocytic vesicles. Among them are Cbl proteins that interact with a CIN85–endophilin complex and control receptor internalization[4]. Previously it has been shown that CIN85 contributes to receptor internalization by clustering Cbl proteins via its three SH3 domains[13]. However, the findings presented here reveal yet another mechanism by which Cbl proteins can contribute to the formation of signal-competent complexes implicated in receptor endocytosis. Biochemical studies in cells clearly indicate that the proline-arginine motif in Cbl can simultaneously accommodate two SH3 domains from separate CIN85 molecules (**Fig. 4**) and could thereby promote the formation of higher-order Cbl–CIN85 complexes. At present we cannot exclude that adjacent SH3 domains in CIN85 may act in a similar manner and cooperate to bind a single ligand. Further studies are required to understand such a potential interplay between SH3A, SH3B and SH3C in the context of the full-length protein. Nevertheless, consistent with our findings, complexes of high molecular mass containing Cbl and CIN85 have been found in cellular lysates[30]. Furthermore, the identification of Arg904 in Cbl-b as an additional binding surface for SH3 domains explains how Cbl-mediated multimerization of CIN85 proteins occurs *in vivo*. The physiological importance of these Cbl functions is likely to be relevant in the assembly of multiprotein complexes during receptor downregulation[13,28,31]. Notably, the β-PIX-SH3 domain binds in the same manner to Cbl proteins as CIN85. Thus, the competition of β-PIX and CIN85 for binding to the same surface of Cbl may represent the basis for dissociation of Cbl away from the activated EGF receptor upon activation of Cdc42–PIX complexes[19].

Taken together, we have identified crucial residues in the proline-arginine motif of c-Cbl and Cbl-b involved in formation of a ternary complex with the adaptor proteins CIN85 and β-PIX and the assembly of the endocytic machinery that governs downregulation of activated receptor tyrosine kinases. Future research should now be directed toward the visualization and dynamics of the formation of Cbl-linked protein complexes in cells and their live imaging during receptor trafficking in the endosome.

METHODS

Protein overexpression and purification. For CIN85A, the N-terminal SH3 domain of human CIN85 (residues 1–58) was cloned into pET29b and overexpressed in *Escherichia coli* strain Rosetta(DE3)pLysS at 30 °C. Cell lysates were precipitated with 50% (w/v) ammonium sulfate, the precipitate was discarded by centrifugation at 35,000*g* for 15 min and more ammonium sulfate was added to the supernatant to 70% saturation. After centrifugation, the CIN85A-containing supernatant was adjusted to 50% (w/v) ammonium sulfate, loaded onto a 5-ml HiTrap Phenyl FF (high sub) hydrophobic interaction column (Amersham Biosciences) and eluted with a linear ammonium sulfate gradient (50% to 0%). CIN85A-containing fractions were pooled, concentrated and loaded onto a HiLoad 16/60 Superdex 75 gel filtration column equilibrated with buffer B (20 mM Tris (pH 7.5) and 150 mM NaCl).

For β-PIX-SH3, the SH3 domain of rat β-PIX (residues 5–66)[15] was cloned into pGEX-4T1 and expressed in *E. coli* strain BL21 at 30 °C. The protein was purified on glutathione Sepharose 4B (Amersham Biosciences) and cleaved on-column with human α-thrombin protease (HTI). Further purification was carried out by gel filtration on Superdex 75 (Amersham Biosciences) in 20 mM Tris (pH 7.5) and 50 mM NaCl. The purified domain was concentrated in VivaSpin concentrators (VivaScience) to ~1 mM and stored at –70 °C after shock freezing in liquid nitrogen.

Isothermal titration calorimetry. Complex formation between the SH3 domains of PIX and CIN85 and various peptides was measured by ITC using a MicroCal VP-ITC microcalorimeter (MicroCal)[32]. The peptides used were Cbl-b, SQAPARPPKPRPRRTAY (the C-terminal tyrosine was introduced to allow spectrophotometric determination of the peptide concentration); and Pak, DATPPPVIAPRPEHTKSVYTR. A NAP 5 column was used to exchange all proteins into ITC buffer (50 mM Tris or HEPES (pH 7.5) and 50 mM NaCl). Experiments were done at 18 °C. In general the sample cell contained either SH3 domain at concentrations of 20–200 μM and peptide solutions at approximately ten times the protein concentration in the cell were injected in 10-μl samples from a total of 290 μl. Heats of dilution, determined by titrating the peptides into buffer alone, were subtracted from the raw titration data before analysis. Data were fitted by least-squares procedures assuming a one-site binding model using Microcal Origin version 7.0 (Microcal). All

Table 2 Data collection and refinement statistics

	β-PIX–Cbl-b	CIN85-SH3A–Cbl-b
Data collection		
Space group	$P3_221$	$P6_5$
Cell dimensions		
a, *b*, *c* (Å)	69.34, 69.34, 58.11	51.16, 51.16, 97.47
Resolution (Å)	26.2–1.85 (1.93–1.85)	20.0–2.0 (2.07–2.00)
R_{sym}	0.086 (0.664)	0.062 (0.430)
$I / \sigma I$	31.8 (2.94)	29.0 (2.87)
Completeness (%)	99.5 (98.1)	99.7 (98.9)
Redundancy	9.7	6.5
Refinement		
Resolution (Å)	26.2–1.85	20.0–2.0
No. reflections	13,340	9,742
R_{work} / R_{free}	22.7 / 26.9	22.1 / 27.1
No. atoms		
Protein	982	931
Ligand/ion	70	94/2
Water	164	82
B-factors		
Protein	29.23	34.99
Ligand/ion	39.91	25.88
Water	42.41	31.92
R.m.s. deviations		
Bond lengths (Å)	0.005	0.007
Bond angles (°)	1.044	1.334

One crystal was used per data set. Values in parentheses are for the highest-resolution shell.

n-Src loops come closest, burying 579 Å^2 of solvent-accessible surface (through direct domain interactions) and contributing to the stability of the SuperSH3 domain conformation through hydrogen bonds across the interface. A GWW motif present in each n-Src loop acts as a signature motif of SuperSH3 domains[27]. The glycine is required for steric complementarity while the second tryptophan residue makes important interactions across the interface. Notably, PIX as well as all three SH3 domains of CIN85 contain this motif (**Fig. 1**). Nevertheless, their relative orientation in the heterotrimeric structures described here is different from that observed in p47^{phox} and residues of the n-Src loop do not contribute to trimerization. This suggests that the new structures presented here are not additional examples of SuperSH3 domains but rather constitute a novel mechanism by which SH3 domains can contribute to the formation of multiprotein complexes.

Cbl-mediated protein multimerization in endocytosis

Efficient receptor endocytosis requires the formation of large protein complexes surrounding activated receptors[4,29]. In many instances, specialized endocytic scaffold proteins engage multiple effectors and couple trafficking cargo with the clathrin coat, actin cytoskeleton and endocytic vesicles. Among them are Cbl proteins that interact with a CIN85–endophilin complex and control receptor internalization[4]. Previously it has been shown that CIN85 contributes to receptor internalization by clustering Cbl proteins via its three SH3 domains[13]. However, the findings presented here reveal yet another mechanism by which Cbl proteins can contribute to the formation of signal-competent complexes implicated in receptor endocytosis. Biochemical studies in cells clearly indicate that the proline-arginine motif in Cbl can simultaneously accommodate two SH3 domains from separate CIN85 molecules (**Fig. 4**) and could thereby promote the formation of higher-order Cbl–CIN85 complexes. At present we cannot exclude that adjacent SH3 domains in CIN85 may act in a similar manner and cooperate to bind a single ligand. Further studies are required to understand such a potential interplay between SH3A, SH3B and SH3C in the context of the full-length protein. Nevertheless, consistent with our findings, complexes of high molecular mass containing Cbl and CIN85 have been found in cellular lysates[30]. Furthermore, the identification of Arg904 in Cbl-b as an additional binding surface for SH3 domains explains how Cbl-mediated multimerization of CIN85 proteins occurs *in vivo*. The physiological importance of these Cbl functions is likely to be relevant in the assembly of multiprotein complexes during receptor downregulation[13,28,31]. Notably, the β-PIX-SH3 domain binds in the same manner to Cbl proteins as CIN85. Thus, the competition of β-PIX and CIN85 for binding to the same surface of Cbl may represent the basis for dissociation of Cbl away from the activated EGF receptor upon activation of Cdc42–PIX complexes[19].

Taken together, we have identified crucial residues in the proline-arginine motif of c-Cbl and Cbl-b involved in formation of a ternary complex with the adaptor proteins CIN85 and β-PIX and the assembly of the endocytic machinery that governs downregulation of activated receptor tyrosine kinases. Future research should now be directed toward the visualization and dynamics of the formation of Cbl-linked protein complexes in cells and their live imaging during receptor trafficking in the endosome.

METHODS

Protein overexpression and purification. For CIN85A, the N-terminal SH3 domain of human CIN85 (residues 1–58) was cloned into pET29b and overexpressed in *Escherichia coli* strain Rosetta(DE3)pLysS at 30 °C. Cell lysates were precipitated with 50% (w/v) ammonium sulfate, the precipitate was discarded by centrifugation at 35,000*g* for 15 min and more ammonium sulfate was added to the supernatant to 70% saturation. After centrifugation, the CIN85A-containing supernatant was adjusted to 50% (w/v) ammonium sulfate, loaded onto a 5-ml HiTrap Phenyl FF (high sub) hydrophobic interaction column (Amersham Biosciences) and eluted with a linear ammonium sulfate gradient (50% to 0%). CIN85A-containing fractions were pooled, concentrated and loaded onto a HiLoad 16/60 Superdex 75 gel filtration column equilibrated with buffer B (20 mM Tris (pH 7.5) and 150 mM NaCl).

For β-PIX-SH3, the SH3 domain of rat β-PIX (residues 5–66)[15] was cloned into pGEX-4T1 and expressed in *E. coli* strain BL21 at 30 °C. The protein was purified on glutathione Sepharose 4B (Amersham Biosciences) and cleaved on-column with human α-thrombin protease (HTI). Further purification was carried out by gel filtration on Superdex 75 (Amersham Biosciences) in 20 mM Tris (pH 7.5) and 50 mM NaCl. The purified domain was concentrated in VivaSpin concentrators (VivaScience) to ~1 mM and stored at –70 °C after shock freezing in liquid nitrogen.

Isothermal titration calorimetry. Complex formation between the SH3 domains of PIX and CIN85 and various peptides was measured by ITC using a MicroCal VP-ITC microcalorimeter (MicroCal)[32]. The peptides used were Cbl-b, SQAPARPPKPRPRRTAY (the C-terminal tyrosine was introduced to allow spectrophotometric determination of the peptide concentration); and Pak, DATPPPVIAPRPEHTKSVYTR. A NAP 5 column was used to exchange all proteins into ITC buffer (50 mM Tris or HEPES (pH 7.5) and 50 mM NaCl). Experiments were done at 18 °C. In general the sample cell contained either SH3 domain at concentrations of 20–200 μM and peptide solutions at approximately ten times the protein concentration in the cell were injected in 10-μl samples from a total of 290 μl. Heats of dilution, determined by titrating the peptides into buffer alone, were subtracted from the raw titration data before analysis. Data were fitted by least-squares procedures assuming a one-site binding model using Microcal Origin version 7.0 (Microcal). All

Table 2 Data collection and refinement statistics

	β-PIX–Cbl-b	CIN85-SH3A–Cbl-b
Data collection		
Space group	$P3_221$	$P6_5$
Cell dimensions		
a, *b*, *c* (Å)	69.34, 69.34, 58.11	51.16, 51.16, 97.47
Resolution (Å)	26.2–1.85 (1.93–1.85)	20.0–2.0 (2.07–2.00)
R_{sym}	0.086 (0.664)	0.062 (0.430)
$I / \sigma I$	31.8 (2.94)	29.0 (2.87)
Completeness (%)	99.5 (98.1)	99.7 (98.9)
Redundancy	9.7	6.5
Refinement		
Resolution (Å)	26.2–1.85	20.0–2.0
No. reflections	13,340	9,742
R_{work} / R_{free}	22.7 / 26.9	22.1 / 27.1
No. atoms		
Protein	982	931
Ligand/ion	70	94/2
Water	164	82
B-factors		
Protein	29.23	34.99
Ligand/ion	39.91	25.88
Water	42.41	31.92
R.m.s. deviations		
Bond lengths (Å)	0.005	0.007
Bond angles (°)	1.044	1.334

One crystal was used per data set. Values in parentheses are for the highest-resolution shell.

29. Dikic, I. Mechanisms controlling EGF receptor endocytosis and degradation. *Biochem. Soc. Trans.* **31**, 1178–1181 (2003).
30. Kowanetz, K. *et al.* CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors. *Mol. Biol. Cell* **15**, 3155–3166 (2004).
31. Watanabe, S. *et al.* Characterization of the CIN85 adaptor protein and identification of components involved in CIN85 complexes. *Biochem. Biophys. Res. Commun.* **278**, 167–174 (2000).
32. Wiseman, T., Williston, S., Brandts, J.F. & Lin, L.N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137 (1989).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
34. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
35. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
36. Perrakis, A., Harkiolaki, M., Wilson, K.S. & Lamzin, V.S. ARP/wARP and molecular replacement. *Acta Crystallogr. D* **57**, 1445–1450 (2001).
37. Turk, D. *Further Development of a Program for Molecular Graphics and Electron Density Manipulation and Its Use in Different Protein Structure Determinations.* PhD Thesis, Technische Universitaet (1992).
38. Laskowski, R.A., Moss, D.S. & Thornton, J.M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).
39. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
40. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
42. McRee, D.E. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
43. Abramoff, M.D., Magelhaes, P.J. & Ram, S.J. Image processing with ImageJ. *Biophotonics International* **11**, 36–42 (2004).

Supplemental Material can be found at:
http://www.jbc.org/cgi/content/full/M606411200/DC1

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 50, pp. 38845–38853, December 15, 2006

Atypical Polyproline Recognition by the CMS N-terminal Src Homology 3 Domain*[S]

Received for publication, July 5, 2006, and in revised form, September 27, 2006 Published, JBC Papers in Press, October 3, 2006, DOI 10.1074/jbc.M606411200

Gabriel Moncalián[‡1], Nayra Cárdenes[‡2], Yonathan Lissanu Deribe[§], Mercedes Spínola-Amilibia[‡], Ivan Dikic[§], and Jerónimo Bravo[‡3]

From the [‡]Signal Transduction Group, Structural Biology and Biocomputing Programme, Centro Nacional de Investigaciones Oncológicas, Melchor Fernández Almagro 3, E-28029 Madrid, Spain and [§]Institute of Biochemistry II, Goethe University Medical School, 60590 Frankfurt, Germany

The CIN85/CMS (human homologs of mouse SH3KBP1/CD2AP) family of endocytic adaptor proteins has the ability to engage multiple effectors and couple cargo trafficking with the cytoskeleton. CIN85 and CMS (Cas ligand with multiple Src homology 3 (SH3) domains) facilitate the formation of large multiprotein complexes required for an efficient internalization of cell surface receptors. It has recently been shown that c-Cbl/Cbl-b could mediate the formation of a ternary complex between one c-Cbl/Cbl-b molecule and two SH3 domains of CIN85, important for the ability of Cbl to promote epidermal growth factor receptor down-regulation. To further investigate whether multimerization is conserved within the family of adaptor proteins, we have solved the crystal structures of the CMS N-terminal SH3 domain-forming complexes with Cbl-b- and CD2-derived peptides. Together with biochemical evidence, the structures support the notion that, despite clear differences in the interaction surface, both Cbl-b and CD2 can mediate multimerization of N-terminal CMS SH3 domains. Detailed analyses on the interacting surfaces also provide the basis for a differential Cbl-b molecular recognition of CMS and CIN85.

Binding of growth factors to receptor tyrosine kinases (RTKs)[4] promotes receptor activation and initiates intracellular signaling cascades that ultimately control biological responses such as cell proliferation, cellular adhesion, or apoptosis (1). Down-regulation of RTKs is a critical step in modulating their activity and is mediated by endocytosis, receptor ubiquitination, and subsequent lysosomal degradation (2). Key to this process is the formation of large protein complexes (3) capable of removing activated receptors from the plasma membrane via clathrin-dependent or -independent pathways (4–7).

The CIN85/CMS (Cas ligand with multiple SH3 domains) family of adaptor proteins participates in the formation of such protein complexes and orchestrates multiple steps in RTK signaling and endocytosis (8). Both CMS (or its mouse ortholog CD2AP) and CIN85 (mouse ortholog SH3KBP1) contain a coiled-coil domain at the C-terminal end, a proline-rich region, and three Src homology 3 (SH3) domains in the N-terminal end. All three domains are responsible for the multiprotein complex formation (8, 9). CIN85 and CMS were suggested to promote clustering of multiple Cbl molecules together with their ability to promote oligomerization through their C-terminal coiled-coil regions. These interactions appear to be important for proper internalization and endocytosis of activated receptor tyrosine kinases (10, 11).

SH3 domains bind proline-rich regions and are essential in the assembly and regulation of many intracellular signaling processes (reviewed in Ref. 12). They are generally composed of ~60 amino acids in a β-barrel structure made up of five β-strands. This arrangement forms one acidic and two hydrophobic pockets between the two β-sheets involved in the recognition and binding of proline-rich motifs (for a review see Refs. 13–15). The majority of SH3 domains characterized to date bind to a proline-rich sequence with a φP$X\varphi$P (PXXP) motif, where φ is usually a hydrophobic residue (16). Each motif adopts a left-handed type II polyproline (PPII) helix. Proline residues are in *trans* conformation, and the helical structure shows precisely three residues/turn (17). Peptide ligands potentially bind a given SH3 domain in either one of two opposite directions governed by the location of a positively charged residue, usually Arg, which often precedes or follows the PXXP core element (18, 19). According to this, the PXXP motif is usually further classified into +$X\varphi$P$X\varphi$P (class I) and φP$X\varphi p$+ (class II) (where + is usually an arginine residue).

Conversely, all CIN85/CMS SH3 domains, can only recognize a novel proline-arginine motif (PXXXPR) present in the C-terminal region of Cbl and Cbl-b (20, 21). Additional cellular proteins, such as CD2 receptor (CD2), Pak, Disabled 1 and 2, SETA-binding protein 1 (SB-1), SLP-65/BLNK, or Alg2-interacting protein (AIP1), and ASAP, contain the consensus sequence PXXXPR and are also bound to CMS/CIN85 in mammalian cells (8, 20). CD2 is a transmembrane cell surface glycoprotein expressed on T lymphocytes, thymocytes, and natural

* This work was supported by Research Grant SAF2003-03860 of the "Ministerio de Ciencia y Tecnología," Spain. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental data.

The atomic coordinates and structure factors (code 2j6f, 2j6o, 2j7i, and 2j6k) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

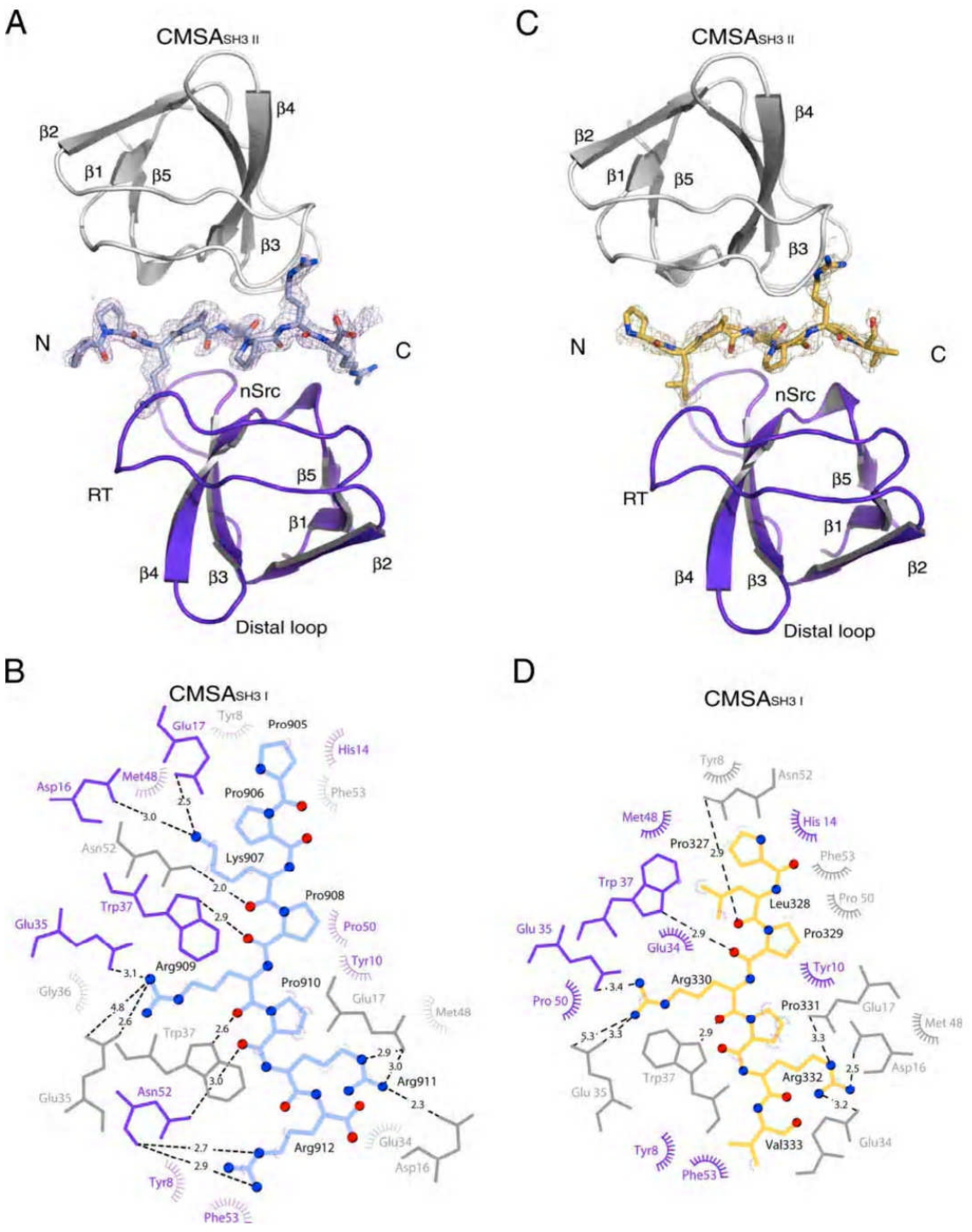
[1] Supported by the "Comunidad de Madrid," Spain. Present address: Departamento de Biología Molecular, Universidad de Cantabria, C/Herrera Oria s/n, 39011 Santander, Spain.

[2] Holder of a fellowship from the "Fundación Ramón Areces," Spain.

[3] To whom correspondence should be addressed. Tel.: 34-912246900 (ext. 2125); Fax: 34-912246976; E-mail: jbravo@cnio.es.

[4] The abbreviations used are: RTK, receptor tyrosine kinase; SH, Src homology; GST, glutathione *S*-transferase; HA, hemagglutinin; wt, wild type; HEK, human embryonic kidney; RT, Arg-Thr.

CMS N-terminal SH3 Domain in Complex with Cbl-b and CD2



killer cells, which mediates low affinity cell/cell interactions by binding to related immunoglobulin superfamily proteins (CD58 in humans). It functions to promote T cell adhesion, immune recognition/activation, and motility (9, 22). Central clustering of CD2 has been proposed to have an important role in maintaining a stable domain for sustained antigen receptor engagement and signaling (23). This CD2 clustering as well as T-cell polarization and recruitment of CD2 into the central contact zone during formation of an immunological synapse has been shown to be dependent on binding of the CMS mouse ortholog CD2AP to the cytoplasmic domain of CD2 (22, 23). The N-terminal SH3 domain of CD2AP appears to have a relatively high specificity and avidity for the conserved proline-rich region of the CD2 cytoplasmic end (22). Both N-terminal CMS and CIN85 SH3 domains (CMSA and CIN85A) have a stronger affinity for CD2 with similar dissociation constants (K_D = ~100 μM) (22) than the rest of CMS and CIN85 SH3 domains (22, 23).

It has recently been shown that Cbl can simultaneously accommodate two SH3 domains from separate CIN85 molecules, promoting the formation of higher order Cbl/CIN85 complexes (24). The study led to the identification of crucial additional residues in the proline-arginine motif of Cbl involved in the formation of a ternary complex with CIN85 N-terminal SH3 and the assembly of the endocytic machinery that governs down-regulation of activated receptor tyrosine kinases. Cbl-b arginines 904 and 911 make equivalent hydrogen bonds with acidic residues in each SH3 domain, allowing the peptide to bind in class I and II simultaneously. The presence of Arg-904 (class I-like binding) defines the multimerization state of the CIN85 SH3-peptide complexes, indicating why other ligands that also contain the P*XXX*PR motif but lack the N-terminal arginine, such as PAK, bind with a 1:1 stoichiometry (24–26). Furthermore, it has also been suggested that other targets of CIN85 that lack the N-terminal arginine, such as BLNK or CD2, should similarly bind in a 1:1 stoichiometry.

CMS is closely related to CIN85; they both interact with Cbl and CD2 and are involved in down-regulation of RTKs (8). To elucidate whether this unique SH3 recognition mechanism is conserved within the CMS/CIN85 family of adaptors and further investigate the role of the N-terminal arginine in the R*XP XXX*PR motif, we have solved the crystal structures of the CMS N-terminal SH3 domain (CMSA) in complex with Cbl-b- and CD2-derived peptides. A differential recognition pattern with respect to CIN85 has been characterized by combining the analysis of structural information and mutagenesis in *in vivo* data. We also aimed at setting up the basis for the atypical proline-rich sequences of Cbl recognition of CMS with respect to CIN85.

EXPERIMENTAL PROCEDURES

Protein Overexpression and Purification—The N-terminal SH3 domain from human CMS (CMSA, residues 1–62) was cloned into pET21 and overexpressed without the His tag in the *Escherichia coli* bv. *rosetta* (DE3) pLys strain at 37 °C. Purification was performed as previously described for CIN85A (24).

Crystallization, Crystallographic Data Collection, Structure Determination, and Refinement—Prior to crystallization, 3 μmol of corresponding peptide were added to 1.2 μmol of purified CMSA protein (70 mg/ml final concentration). All crystals were obtained at 20 °C using the hanging drop vapor diffusion method by mixing 1 μl of protein solution with 1 μl of reservoir solution. Crystals were then transferred to stabilization buffer (10% (v/v) ethylene glycol, 90% of reservoir solution) and flash-cooled to 110 °K. CMSA/Cbl-b (Cbl-b PARPPKPRPRR; Genosys) was crystallized in the space group I422 with cell dimensions a = 66.51, b = 66.51, and c = 67.78 Å of CMSA:Cbl-b (Table 1) with reservoir solution containing 20% polyethylene glycol 3000, 0.1 M acetate at pH 5.5.

The CMSA/CD2 heterodimer (CD2 KGPPLPRPRV; Genosys) was crystallized with reservoir solution of 30% polyethylene glycol 8000, 0.2 M NaCl, 0.1 M Tris-HCl, pH 8.0, in the space group $P2_12_12_1$ with cell dimensions a = 36.37, b = 59.80, and c = 71.12 Å and the CMSA/CD2 heterotrimer crystallized in the I422 space group with cell dimensions a = 66.394, b = 66.448, and c = 68.739 Å (crystallization solution: 20% polyethylene glycol 3000, 0.1 M acetate, pH 5.5). All data sets were collected in a Mar 345 imaging plate and a Bruker rotating anode source and were processed with DENZO and SCALEPACK software (27) for CMSA/Cbl-b or MOSFLM and SCALA software (28) for CMSA/CD2 complexes.

The CMSA:Cbl-b complex structure was solved by molecular replacement using AMoRe software (29) and the coordinates of an isolated CIN85A as the search model.[5] Similarly, the CMSA/Cbl-b structure was used as a template for solving CMSA/CD2 structures. The structures were refined by cycles of positional and restrained individual B-value refinement with CNS (30) and Refmac5 software (28) and by manual model building using Xfit (31) or Coot software (32).

Biochemical Assays —Cell transfection with the indicated cDNAs, cell lysis, glutathione *S*-transferase (GST) pulldowns, immunoprecipitations, and immunoblotting were essentially performed as described previously (20).

RESULTS

CMSA/Cbl-b Overall Crystal Structure—To understand the molecular details of the different members of the CIN85/CMS family of adaptor molecules, we have crystallized the N-termi-

[5] N. Cárdenes, G. Moncalián, and J. Bravo, unpublished data.

FIGURE 1. **Overall structure of the CMSA·Cbl-b and CMSA·CD2 heterotrimeric complexes and protein/peptide interaction details.** *A* and *C*, $CMSA_{SH3I}$ binds the peptides in a class I orientation (shown in *purple*), and $CMSA_{SH3II}$ recognizes the peptide in class II orientation (shown in *gray*). The Cbl-b peptide is shown in *light blue* (*A*) and CD2 peptide in yellow (*B*). The σA-weighted electron density map around them is contoured at 1.0 σ. Elements of secondary structure and the positions of the RT, *n*-Src, and distal loops are indicated. *B* and *D*, Schematic representation of contacts between CMSA and the Cbl-b (*B*) and CD2 (*D*) peptides (*light blue* and *yellow*, respectively). Residues 902, 903, and 904 from the Cbl-b peptide were not visible at the electron density map. Residues 324, 325, and 326 from the CD2 peptide are also disordered. Interactions and residues from SH3I are shown in *purple* and in *gray* for SH3II. *Dashed lines* show hydrogen bonds (labeled with peptide-protein distances in Å), and *purple* and *gray rays* designate hydrophobic interactions.

nal SH3 domain of CMS (CMSA) in complex with the same Cbl-b-derived peptide used for the crystallization of the CIN85A/Cbl-b complex (24).

The refined structure at 1.7 Å of CMSA/Cbl-b shows a ternary complex composed of two CMSA molecules binding, in opposite orientations, to a single Cbl-b peptide molecule (Fig. 1*A*). One SH3 domain recognizes Lys-907 in the N terminus of the peptide, resembling a classical class I orientation, and the second SH3 domain recognizes Arg-911 in the C terminus, resembling class II. The asymmetric unit contains one CMSA molecule bound to a single Cbl-b peptide, which lies along a crystallographic 2-fold axis. Thus, a crystallographically related CMSA molecule binds the peptide on its opposite side, and electron density maps show an average of two orientations for the peptide (see supplemental figure). Final refined B factors for the peptide, considering occupancy of 0.5 for each orientation, do not differ significantly from the B factors of the CMSA-interacting residues (~25 Å^2). The double orientation of the peptide is possible, because crystallographic contacts are exclusively established by CMSA molecules. CMSA residues 1 and 59–62 are not visible in the electron density maps. Cbl-b residues 902, 903, and 904 were also disordered.

The structure of the CMSA molecule displays the typical SH3 domain folding, adopting the characteristic five-stranded β-barrel structure. The peptide, on the other hand, does not display a typical polyproline helix but adopts an extended conformation instead. The structure shows that CMS also has the ability to form heterotrimeric complexes with a Cbl-b derived peptide, as it has recently been reported for CIN85 (24).

CMSA/Cbl-b Molecular Interface—In the heterotrimer arrangement, the interactions of the $CMSA_{SH3I}$ (class I-like orientation) and $CMSA_{SH3II}$ (class II-like orientation) with the Cbl-b peptide are almost identical with 941 and 827 Å^2 of buried surface, respectively, and are comparable with values observed for other SH3 domain/peptide structures. However, the interacting surface of the two CMSA molecules is negligible. Only 54 Å^2 of the 3918 Å^2 total surface of each SH3 is shared, corresponding mainly to both side chains of Glu-35 (from $CMSA_{SH3I}$ and $CMSA_{SH3II}$).

Hydrophilic specific interactions via side chains are only made by Lys-907 and Arg-911 at the N and C termini of the Cbl-b peptide with both Asp-16 and Glu-17 from $CMSA_{SH3I}$ and $CMSA_{SH3II}$ respectively and Arg909 with Glu-35 from both CMSA molecules (Fig. 1*B*). Hydrophobic contacts in the class I orientation are made by Cbl-b Pro-910 with conserved residues of hydrophobic pocket II. In turn, Cbl-b Pro-906 and -908 are embedded within hydrophobic pockets I and II, respectively. Thus, Cbl-b Pro-908 and -910 occupy equivalent positions in the opposed SH3 domains of the heterotrimer. Additionally, there are sequence-unspecific contacts with peptide main chain atoms, such as Pro-910 and Lys-907 carboxylic oxygens interacting, respectively, with Asn-52 from each $CMSA_{SH3I}$ and $CMSA_{SH3II}$. Similarly, Pro-908 and Arg-909 carboxylic oxygens interact with Trp-37 from both $CMSA_{SH3I}$ and $CMSA_{SH3II}$ (Fig. 1*B*).

In Vivo Interactions between Cbl-b and CMS—The structures of the CMSA domain and the peptide derived from Cbl-b containing the P*XXX*PR motif shows that Cbl-b Lys-907 and Arg-911 residues are important for the interaction between the SH3 domain and its ligand (Fig. 1*B*).

We set out to see the importance of these residues for the interaction between Cbl-b and CMS in the context of full-length proteins expressed in cells. To this end, expression vectors encoding HA-tagged Cbl-b wt and Cbl-b K907A, Cbl-b R911A, and Cbl-b K907A/R911A point mutants were transfected and overexpressed in HEK293T cells. Lysates from these cells were then used to test binding to GST-CMSA in GST pulldown assays. Fig. 2*A* shows that both Cbl-b Lys-907 and Arg-911 play important roles in the interaction, although the Cbl-b K907A mutation abrogated binding to a much lesser extent than Arg-911. Double mutation of both Cbl-b Lys-907 and Arg-911 abolished the binding almost completely.

Because of the overall similarity in the P*XXX*PR motif in Cbl-b and c-Cbl, we then checked whether the corresponding residues in c-Cbl also mediate the interaction with CMSA. Thus, we mutated the equivalent residues Lys-825 and Arg-829 to alanines in c-Cbl and proceeded as for GST pulldown assays with Cbl-b transfections. Interestingly, a very similar profile of binding to Cbl-b was observed. Again both c-Cbl Lys-825 and c-Cbl Arg-829 appear to be important for binding to the CMSA domain. The c-Cbl K825A mutation impaired binding to the SH3 domain slightly, whereas R829A led to a much greater loss in binding, and the double point mutations largely abrogated the interaction of c-Cbl with the N-terminal SH3 domain of CMS (Fig. 2*B*). Both experiments reproducibly showed the relative importance of these residues to the binding between the PK*XX*PR motif in Cbl-b/c-Cbl and CMSA.

Further on, we were interested to see the roles of Cbl-b Lys-907 and Cbl-b Arg-911, and the corresponding residues in c-Cbl in the binding of full-length Cbl-b/c-Cbl and CMS. To this end, expression vectors encoding HA-tagged Cbl-b wt and K907A, R911A, and K907A/R911A mutants along with FLAG-tagged CMS were transfected in HEK293T cells. Cell lysates were subjected to immunoprecipitation with an anti-Cbl-RF antibody (recognizing the ring finger region of both Cbl-b and c-Cbl) and immunoblotted with anti-FLAG antibody.

As shown in Fig. 2*C*, FLAG-CMS is co-precipitated quite strongly with wt Cbl-b. This is in sharp contrast to Cbl-b K907A and R911A mutants, where a minimal amount of FLAG-CMS is co-immunoprecipitated with the mutants. This result demonstrates the vital importance of these residues in binding to CMS. The residual binding seen in the case of the Cbl-b K907A/R911A double mutant is most likely because of FLAG-CMS being co-precipitated with endogenous Cbl-b/c-Cbl that is immunoprecipitated by the Cbl-RF antibody.

The same assay was performed with the transfection of HA-c-Cbl wt, K825A, R829A, and K825A/R829A point mutants together with FLAG-CMS. It can be clearly seen that FLAG-CMS is co-immunoprecipitated with wt c-Cbl; however, the c-Cbl K825A and/or R829A mutation markedly abolished the co-immunoprecipitation of CMS (Fig. 2*C*, *lower panel*). Thus, in a similar fashion to Cbl-b, c-Cbl K825A and R829A mutants seem to be important for binding to CMS.

The crystal structure of CMSA:Cbl-b suggests a different interaction pattern in class I recognition compared with CIN85A/Cbl-b. Cbl-b Lys-907 in CMSA/Cbl-b occupies an

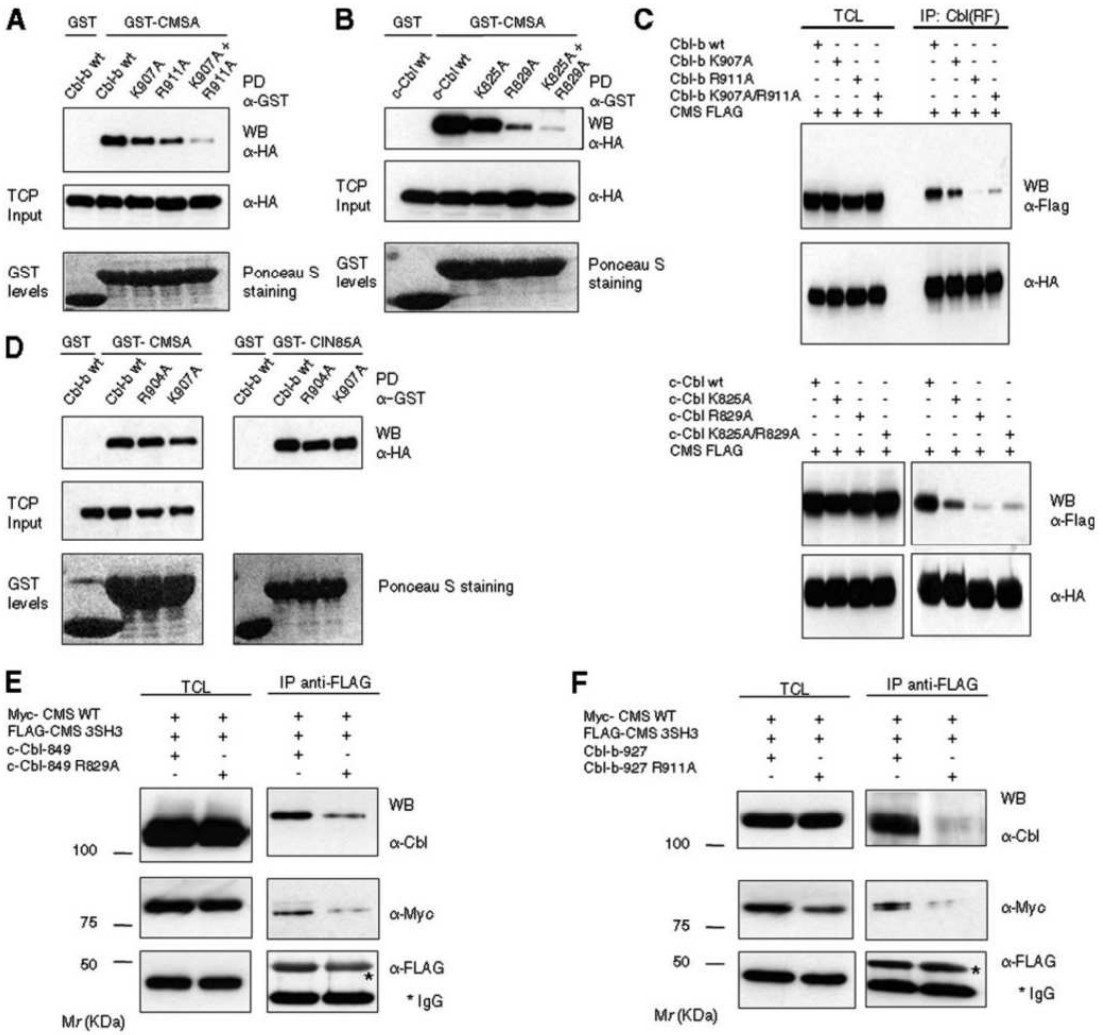


FIGURE 2. **Lys-907 and Arg-911 from Cbl-b play a crucial role in CMSA/Cbl-b binding and heterotrimerization.** *A*, HEK293T cells were transfected with HA-Cbl-b wt, HA-Cbl-b K907A, HA-Cbl-b R911A, or HA-Cbl-b K907A/R911A. Lysates were incubated with GST alone or GST-CMSA, and the inputs and pulldowns (*PD*) were then subjected to immunoblotting (*WB*) with anti-HA antibody. *B*, GST pulldowns and WB were performed also with lysates of HEK293T cells transfected with HA-c-Cbl K825A, R829A, or K825A/R829A point mutants. *C*, HA-tagged full-length Cbl-b wt, K907A, R911A, K907A/R911A mutants and c-Cbl wt, K825A, R829A, and R825A/R829A mutants were transfected in HEK293T cells along with FLAG-tagged full-length CMS. Cell lysates were subjected to immunoprecipitation with an anti-Cbl-RF antibody (recognizing the ring finger region of both Cbl-b and c-Cbl) and immunoblotted with anti-FLAG antibody. *D*, comparison of Cbl-b Arg-904 *versus* Lys-907 in CMSA/Cbl-b and CIN85A/Cbl-b binding. HEK293T cells were transfected with HA-Cbl-b wt, HA-Cbl-b R904A, or HA-Cbl-b K907A. Lysates were incubated with GST alone, GST-CMSA, or GST-CIN85A. The inputs and pulldowns were then subjected to immunoblotting (*WB*) with anti-HA antibody. Shown are HEK293T cells transfected with FLAG-CMSA-3SH3, full-length Myc-tagged CMS, and c-Cbl-849 (*E*) or Cbl-b-927 (*F*) constructs lysed and immunoprecipitated by anti-FLAG affinity agarose beads. Total cell lysates (*TCL*) and immunoprecipitates (*IP*) were subjected to immunoblotting (*WB*) with corresponding antibodies.

equivalent position as Arg-904 in the CIN85A/Cbl-b structure (see Figs. 1, *A* and *B*; and 5). Therefore, we wanted to compare the relative importance of Cbl-b Arg-904 and Cbl-b Lys-907 residues in binding to N-terminal SH3 domains from CMS and CIN85. Fig. 2*D* shows that GST-CMSA was able to bring down Cbl-b R904A as potently as wild type but pulled down Cbl-b K907A only weakly. In sharp contrast, GST-CIN85A was able to pull down both wild type Cbl-b and Cbl-b K907A mutant equally potently, with a slight reduction with respect to Cbl-b R904A in accordance with a previous observation (24).

The structure of CMSA/Cbl-b shows the formation of heterotrimers involving two CMS N-terminal SH3 domains and one Cbl peptide *in vitro*. To check whether the formation of CMSA/Cbl heterotrimers can also take place *in vivo*, HEK293T cells were transfected with a FLAG-tagged construct of CMS containing only the three SH3 domain regions, Myc-tagged

full-length CMS, and either c-Cbl-849 or Cbl-b-927 constructs that lack the ability to homodimerize. These Cbl mutants were used to ensure that co-precipitation between CMS molecules was a consequence of direct binding of two independent CMS molecules to a single Cbl molecule rather than due to the interaction with homodimeric Cbl molecules (24). Cells overexpressing c-Cbl-849 or Cbl-b-927 have the ability to co-precipitate full-length CMS using a FLAG antibody (Fig. 2, *E* and *F*), indicating that Cbl is mediating the formation of the heterotrimer in cells. We have shown that point mutants of c-Cbl or Cbl-b are significantly impaired in binding to CMS (Fig. 2, *A–D*). When checking for heterotrimer formation in cells using the same point mutations in the constructs that lack the ability to homodimerize (Cbl-b-927 R911A and c-Cbl-849 R829A), the amount of co-precipitated full-length CMS is markedly diminished (Fig. 2, *E* and *F*, respectively). It is important to note that, because we cannot avoid the endogenous Cbl proteins that can mediate the formation of a heterotrimeric complex, we have detected some full-length CMS co-precipitated even in the absence of overexpressed Cbl proteins. Taken together, we can conclude that a complex containing Cbl proteins and two intermolecular SH3 domains of CMS can be formed in cells in the context of full-length CMS proteins.

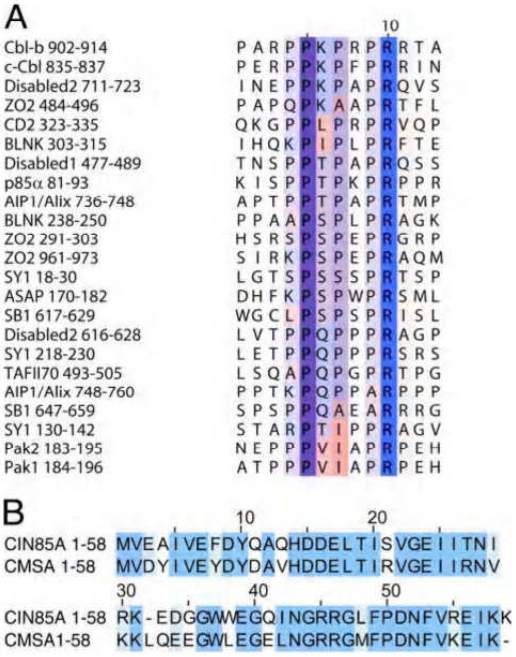


FIGURE 3. **Alignment of interacting sequences from CIN85/CMS partners and of CMSA and CIN85A.** *A*, several polyproline-rich sequences containing the P*XXX*PR motif from CIN85/CMS interacting molecules were aligned. Hydrophobic residues are shown in conservation gradient from *blue* to *red*. *B*, sequence alignment of the N-terminal SH3 domains from CMS and CIN85 was performed. Identical residues are shown in *blue* and conserved hydrophobic residues in *light blue*.

CMSA/CD2 Crystal Structures—Structural and *in vivo* CMSA/Cbl-b results demonstrate the ability of CMS to form heterotrimeric complexes with Cbl-b-derived peptides in a similar fashion as was reported for CIN85. We wanted to check whether a similar CMS recognition was possible for other atypical polyproline-containing molecules. We observed, in the CMSA/Cbl-b structure, that Lys-907 is involved in the interaction with SH3I (Fig. 1*B*). However, this residue is only conserved in Cbl-b or c-Cbl proteins, and it is not present in some other CMS/CIN85-interacting proteins such as CD2, SB-1, SLP-65/BLNK, or AIP1 (Fig. 3*A*). In CD2, the position of the lysine is occupied by a leucine (PLPRPR). The fact that the Cbl-b Lys-907 side chain interacts with the RT loop of SH3I in the CMSA/Cbl-b crystal structure (Fig. 1*D*) could suggest that a CD2-derived peptide could be predominantly a class II-like interaction. To check this possibility, the crystal structure of the CMSA/CD2-derived peptide complex was solved at 2.9 and 2.2 Å (see "Experimental Procedures" and Table 1). Two different crystals were obtained. Unexpectedly, one of the structures was a heterotrimer of two SH3 domains binding to the peptide in the same mode that Cbl-b-derived peptide is being recognized (Fig. 1*C*), yet the other structure is of a single SH3 domain recognizing the peptide in a class II orientation. Both CMSA/CD2 and CMSA/Cbl-b heterotrimeric structures display similar peptide recognition modes, with the exception of contacts established by Cbl-b Lys-907 (Leu-328 in CD2) (Fig. 1*D*).

TABLE 1
Crystallographic data collection and analysis

	CMSA/Cbl-b	CMSA/CD2 dimer	CMSA/CD2 trimer
Data collection			
Space group	I422	$P2_12_12_1$	I422
Unit cell dimensions			
a, *b*, *c* (Å)	66.51, 66.51, 67.78	36.32, 59.72, 70.87	66.43, 66.43, 68.72
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Data range (Å)	20–1.7	36.3–2.9	34.4–2.5
Observations (unique)	107,249 (8,696)	139,722 (3,729)	66,722 (4,006)
Completeness (%) (last shell)	99.9 (100.0)	100.0 (100.0)	98.0 (86.6)
R_{sym}[a] (last shell)	0.034 (0.269)	0.095 (0.307)	0.110 (0.408)
Mean I/σ (last shell)	20.9 (6.9)	22.0 (6.2)	24.5 (7.5)
Refinement			
Non-hydrogen atoms (solvent molecules)	641 (60)	1104 (32)	595 (69)
R_{work}[b] (R_{free})[c]	0.189 (0.218)	0.245 (0.293)	0.241 (0.286)
Root mean square bond length (Å)	0.012	0.0080	0.010
Root mean square bond angles (°)	1.379	1.400	1.306

[a] R_{sym} is the unweighted *R* value on *I* between symmetry mates.
[b] $R_{work} = \Sigma_{hkl}[[F_{obs}(hkl)] - [F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$.
[c] R_{free} is the cross-validation *R*-factor for 5% of reflections against which the model was not refined.

DISCUSSION

CMSA/Cbl-b Forms a Heterotrimeric Complex with a Differential Recognition from CIN85A/Cbl-b—Both CMS and CIN85 share the same domain organization with a 39% sequence identity (~70%) within the conserved domains (Fig. 3*B*). This similarity suggests that they may have overlapping binding specificities. Consequently, they have been assumed to share similar biological functions (8). Superposition of CMSA/Cbl-b and CIN85A/Cbl-b structures reveal that the overall structure of CMSA/Cbl-b resembles the heterotrimer reported for CIN85A/Cbl-b (24). One SH3 domain binds the Cbl-b peptide in a class I-oriented fashion, whereas the second SH3 recognizes the peptide in the opposite orientation (class II). Such similarity in the overall structures of the heterotrimeric complexes provides a molecular explanation for a common role of Cbl in the molecular mechanism of CIN85/CMS-mediated down-regulation of RTKs. Cbl might also participate in the clustering of CMS as reported for CIN85 and β-PIX (24). The C-α of the SH3 individual domains overlap well with one another (root mean square deviation, 0.70 Å) (Fig. 4). Main differences are located within the *n*-Src and distal loops. Nevertheless, the overall CMSA/Cbl-b heterotrimer shows some important differences with respect to CIN85A/Cbl-b. In particular, the orientation of the SH3 domains in CMSA/Cbl-b differs from that of the SH3 orientations in CIN85A/Cbl-b, revealing a distinctive recognition of the Cbl-b-derived peptide for CMS or CIN85 (Fig. 4).

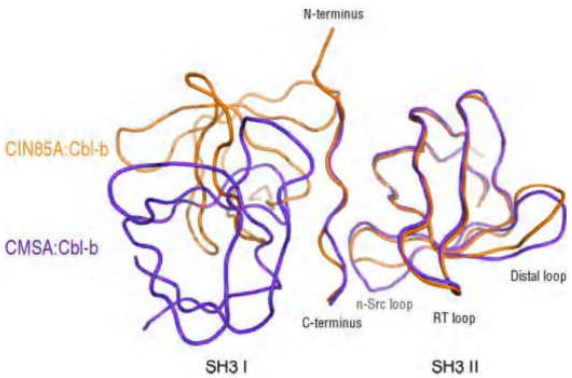


FIGURE 4. **Comparison of the ternary CMSA/Cbl-b complex with the reported CIN85A/Cbl-b complex (2BZ8).** SH3II backbones from the CMSA/Cbl-b (*purple*) and CIN85A/Cbl-b (*orange*) complexes are overlaid. *n*-Src, RT, and distal loops are indicated.

The interactions corresponding to the class II orientation of the peptide ($CMSA_{SH3II}$/Cbl-b) are conserved with respect to CIN85 ($CIN85A_{SH3II}$/Cbl-b) (Fig. 1*B*). Arg-911 determines the class II orientation, and its importance in (CMS/CIN85)/Cbl interaction has already been described and proven for CIN85 (20, 21). The conserved heterotrimeric arrangement suggests it also plays a critical role in CMS/Cbl-b class II interaction. The only difference in class II recognition between both structures concerns Cbl-b Arg-909, which hydrogen bonds with Asp-33 from CIN85 or Glu-35 from CMS (Fig. 1*B*).

However, significant differences in the interaction surface between Cbl-b and CMSA or CIN85A are observed in the class I orientation. Residues forming the two hydrophobic binding pockets are highly conserved between CMSA and CIN85A, yet the peptide residues interacting with their SH3I domains are different. Most differences are found in the acidic pocket formed by the *n*-Src and RT loops. Comparison of the uncomplexed CMSA (supplemental data) and CIN85A crystal structures show no conformational changes in the *n*-Src and RT loops upon peptide binding (root mean square deviations, 0.42 and 0.64 Å, respectively), and therefore, differences observed between CMSA/Cbl-b and CIN85/Cbl-b are not due to conformational changes upon peptide binding.

When compared with the CIN85A/Cbl-b-reported structure, the N terminus of the Cbl-b peptide adopts a different conformation that mostly reflects changes in the class I binding orientation. A significant three-residue shift of the Cbl-b peptide is observed when comparing $CMSA_{SH3I}$ and $CIN85_{SH3I}$ (Fig. 5). Resultant $CMSA_{SH3I}$/Cbl-b interactions include hydrogen bonds of Asp-16 and Glu-17 side chains to Lys-907, contacts of the Pro-908 main chain with the conserved Trp-37, and hydrophobic interactions of the Pro-910 side chain with residues from hydrophobic pocket II.

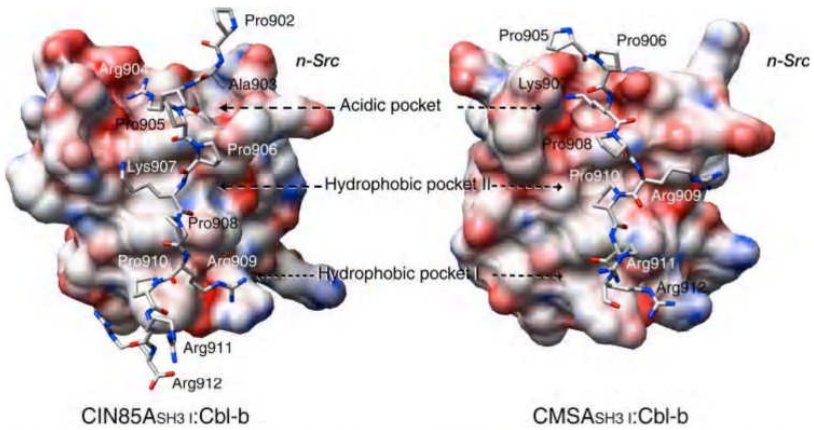


FIGURE 5. **Comparison of the SH3I/Cbl-b interface from CIN85A and CMSA.** Electrostatic potential surfaces of SH3 domains from CIN85A/Cbl-b (*left*) and CMSA/Cbl-b (*right*) complexes that bind the peptide in a class I-like orientation are shown. The Cbl-b peptides are represented in *sticks*, and their residues are indicated. *n*-Src loops are labeled in both structures, and binding pockets are indicated with *arrows*.

In CIN85A, Asp-16 and Glu-17 from the RT loop hydrogen bond with Cbl-b Arg-904, which plays a critical role in the formation of the CIN85A/Cbl-b heterotrimeric complex. However, in CMSA/Cbl-b, Arg-904 does not participate in the interaction, and it is the residue Lys-907 that interacts with the RT loop in the acidic pocket (Fig. 5).

In addition, Cbl-b Arg-909 in CIN85A/Cbl-b occupies a similar position to Cbl-b Arg-912 in the CMSA/Cbl-b structure, and hydrophobic pocket I is empty in $CMSA_{SH3I}$, because the synthesized peptide ends at Arg-912. The next

residue in the Cbl-b sequence is Thr-913, and there is certainly room for its accommodation in CMSA/Cbl-b. Overall, the three-residue shift observed for class I recognition leads to the conclusion that the Cbl-b sequence interacting with CIN85 might differ from the optimal Cbl-b sequence interacting with CMS, affecting the affinity for both adaptor molecules. This could have severe consequences in the multiprotein complex formation responsible for RTK down-regulation and other cellular processes.

This three-residue shift in the class I orientation can be rationalized by sequence and structural differences within the respective *n*-Src loops (Figs. 3*B* and 4), which have been reported to account for binding selectivity (16). With respect to the CIN85 sequence, a leucine is inserted in CMS at position 32 that significantly changes the overall conformation of the *n*-Src loop. In addition, Glu-35 (Gly-34 in CIN85) hydrogen bonds Arg-909 from the peptide. The N-terminal end of the peptide faces opposite directions in each structure, and such displacement might be due to the change in the conformation of the *n*-Src loop, given that Glu-34 in CMSA does not allow the peptide to adopt a conformation similar to the one displayed in the CIN85A·Cbl-b complex (Fig. 5).

CMSA Interacts with CD2 Forming Two Types of Molecular Arrangements—We have shown that residue Lys-907 from Cbl-b plays an important role on CMSA binding (Figs. 1*B* and 2, *A* and *C*). Cbl-b Lys-907 is not conserved for other CMS ligands, and CD2 contains Leu-328 at the equivalent position. This would suggest that at least the class I type of interaction could be compromised upon CD2 binding to CMSA. In fact, we have obtained crystals of a dimeric form of a 1:1 CMSA·CD2 complex on a class II type of interaction. Unexpectedly, an additional heterotrimeric crystal has also been obtained in an independent crystal form. The existence of the two molecular arrangements suggests a minimal energy difference between the heterotrimeric and dimeric states that might have some biological consequences in contrast to the predominant heterotrimeric complex formed by CMSA/Cbl-b.

In the CMSA·Cbl-b complex, Cbl-b Lys-907 hydrogen bonds to CMS Asp-16 and Glu-17. This interaction cannot take place with CD2 because of a substitution of the equivalent Cbl-b lysine to leucine. The total non-polar buried surface upon CMSA binding is similar for Cbl-b Lys-907 and CD2 Leu-328 (101 and 120 Å^2, respectively); thus, the main difference is the absence of the Lys-907-mediated hydrogen bond. However, the absence of this hydrogen bond is not enough to prevent the formation of the CMSA/CD2 heterotrimer.

In addition, CD2 peptides obtained in the heterotrimer and dimeric forms perfectly superimpose (root mean square deviation, 0.48 Å). The two independent CMSA/CD2 structures present entirely different crystal packing, indicating that the extended conformation of the peptide observed in CMSA·peptide complex structures cannot be due to packing forces.

CIN85 and CMS SH3 Binding Consensus Motifs—Biochemical experiments shown here provide *in vivo* confirmations for the observed structures and the residues involved in the binding between the SH3 domains and their ligands. GST pulldown and co-immunoprecipitation experiments shown in Fig. 2 clearly indicate that Cbl-b Lys-907 and Cbl-b Arg-911 and the corresponding residues in c-Cbl play an important role in binding to CMS in the context of full-length proteins in cells. Furthermore, we have shown a differential binding of SH3 domains of CMS and CIN85 to Cbl-b. This is in such a way that Cbl-b Arg-904 is more important for binding to CIN85A than to CMSA, whereas the opposite holds true for Cbl-b Lys-907, which is observed to be more important for binding to CMSA. In addition, we have also shown that Cbl mediates the formation of heterotrimers in the context of full-length CMS. Cbl-mediated multimerization of CMS proteins occurring *in vivo* suggests similar roles between CMS and CIN85. Despite the differential Cbl molecular recognition between the two adaptor molecules, they both contribute to the Cbl-mediated clustering and formation of large protein complexes required for receptor down-regulation.

The structures presented here unambiguously indicate that residues involved in class I recognition are different from those involved in class II. Cbl-b Lys-907 and Pro-910 are responsible for CMSA/Cbl-b class I orientation. Specific key residues such as Cbl-b Arg-904 might only be essential for CIN85A heterotrimerization and not for CMSA heterotrimerization. Besides, we have shown, from the CMSA/CD2 interaction, that Cbl-b Lys-907 can be replaced by a Leu and still form a heterotrimer. Given that hydrophobic pocket I is empty in $CMSA_{SH3I}$/Cbl-b and $CMSA_{SH3I}$/CD2, the motif could be extended to $\varphi XXPXX$ for CMS class I and $PXPXXR$ for both CIN85 and CMS class II. Taken together, our results indicate that consensus sequences should indeed be considered independent for class I and class II recognition types. Nevertheless, the overall motif is similar to the one described by Kowanetz *et al.* (20) ($PXP/AXPR$), and a structure of CMSA/Cbl-b P908A at 1.9 Å resolution confirms that the same heterotrimerization is possible with Pro-908 replaced by an alanine (supplemental data). It has been shown that some other hydrophobic residues could be present at this position without significantly affecting the interaction, such as isoleucine as found in the Pak polyproline motif (Fig. 3*A*) (26). Thus, a more specific CMS consensus motif could be narrowed down to $P\varphi(P/A)XPRX$ for the heterotrimer formation and $RXPX(P/A)XPR$ in the case of CIN85.

The process of endocytosis requires the formation of large multiprotein complexes for efficient internalization of cell surface proteins. To a large extent, this is facilitated by the so-called endocytic adaptor proteins that could engage multiple effectors and couple cargo trafficking with the cytoskeletal apparatus (3). The CIN85/CMS family of proteins represents an example of such adaptors. These proteins enhance endocytosis by clustering Cbl proteins through their three SH3 domains (20). The existence of yet another level of organization is also suggested by the recent observation that c-Cbl/Cbl-b proteins can mediate the formation of a ternary complex between one c-Cbl/Cbl-b molecule and two SH3 domains of CIN85 (24). The structural and biochemical study provided here further supports the notion of additional clustering via SH3 domain dimerization through a c-Cbl/Cbl-b molecule and extends a similar possibility for CD2.

Acknowledgments—We thank Miguel Ortiz for help in the refinement and critical review of the manuscript and José Rivera and Carme Fábrega for useful discussion.

REFERENCES

1. Schlessinger, J. (2002) *Cell* **110,** 669–672
2. Waterman, H., and Yarden, Y. (2001) *FEBS Lett.* **490,** 142–152
3. Dikic, I., and Giordano, S. (2003) *Curr. Opin. Cell Biol.* **15,** 128–135
4. Bonifacino, J. S., and Traub, L. M. (2003) *Annu. Rev. Biochem.* **72,** 395–447
5. Szymkiewicz, I., Shupliakov, O., and Dikic, I. (2004) *Biochem. J.* **383,** 1–11
6. Sigismund, S., Woelk, T., Puri, C., Maspero, E., Tacchetti, C., Transidico, P., Di Fiore, P. P., and Polo, S. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 2760–2765
7. Le Roy, C., and Wrana, J. L. (2005) *Nat. Rev. Mol. Cell Biol.* **6,** 112–126
8. Dikic, I. (2002) *FEBS Lett.* **529,** 110–115
9. Tibaldi, E. V., and Reinherz, E. L. (2003) *Int. Immunol.* **15,** 313–329
10. Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W. Y., and Dikic, I. (2002) *Nature* **416,** 183–187
11. Szymkiewicz, I., Kowanetz, K., Soubeyran, P., Dinarina, A., Lipkowitz, S., and Dikic, I. (2002) *J. Biol. Chem.* **277,** 39666–39672
12. Zarrinpar, A., Bhattacharyya, R. P., and Lim, W. A. (2003) *Science's STKE* 2003, RE8
13. Cesareni, G., Panni, S., Nardelli, G., and Castagnoli, L. (2002) *FEBS Lett.* **513,** 38–44
14. Larson, S. M., and Davidson, A. R. (2000) *Protein Sci.* **9,** 2170–2180
15. Ferreon, J. C., and Hilser, V. J. (2004) *Biochemistry* **43,** 7787–7797
16. Mayer, B. J. (2001) *J. Cell Sci.* **114,** 1253–1263
17. Adzhubei, A. A., and Sternberg, M. J. (1993) *J. Mol. Biol.* **229,** 472–493
18. Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994) *Science* **266,** 1241–1247
19. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) *Nature* **372,** 375–379
20. Kowanetz, K., Szymkiewicz, I., Haglund, K., Kowanetz, M., Husnjak, K., Taylor, J. D., Soubeyran, P., Engstrom, U., Ladbury, J. E., and Dikic, I. (2003) *J. Biol. Chem.* **278,** 39735–39746
21. Kurakin, A. V., Wu, S., and Bredesen, D. E. (2003) *J. Biol. Chem.* **278,** 34102–34109
22. Hutchings, N. J., Clarkson, N., Chalkley, R., Barclay, A. N., and Brown, M. H. (2003) *J. Biol. Chem.* **278,** 22396–22403
23. Dustin, M. L., Olszowy, M. W., Holdorf, A. D., Li, J., Bromley, S., Desai, N., Widder, P., Rosenberger, F., van der Merwe, P. A., Allen, P. M., and Shaw, A. S. (1998) *Cell* **94,** 667–677
24. Jozic, D., Cardenes, N., Deribe, Y. L., Moncalian, G., Hoeller, D., Groemping, Y., Dikic, I., Rittinger, K., and Bravo, J. (2005) *Nat. Struct. Mol. Biol.* **12,** 972–979
25. Mott, H. R., Nietlispach, D., Evetts, K. A., and Owen, D. (2005) *Biochemistry* **44,** 10977–10983
26. Hoelz, A., Janz, J. M., Lawrie, S. D., Corwin, B., Lee, A., and Sakmar, T. P. (2006) *J. Mol. Biol.* **358,** 509–522
27. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
28. Collaborative Computational Project (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
29. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50,** 157–163
30. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
31. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165
32. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132

CMS N-terminal SH3 domain in complex with Cbl-b and CD2

SUPPLEMENTAL DATA

Supplemental Table I. Crystallographic data collection and analysis

	CMSA	CMSA:Cbl-b P908A
Data collection		
Space group	I4	I422
Unit Cell Dimensions		
a, b, c (Å)	120.02, 120.02, 153.93	67.44, 67.44, 69.48
α, β, γ (°)	90, 90, 90	90, 90, 90
Data range (Å)	47.30-2.74	20.0-1.9
Observations (unique)	95948 (27678)	47652 (6606)
Completeness (%) (last shell)	97.2 (81.0)	99.9 (100.0)
R_{sym}[a] (last shell)	0.116 (0.388)	0.047 (0.330)
Mean I/σ (last shell)	10.9 (3.6)	17.4 (4.5)
Refinement		
Non-hydrogen atoms (solvent molecules)	5889 (103)	575 (30)
R_{work}[b] (R_{free}[c])	0.207 (0.272)	0.220 (0.256)
R.m.s. bond length (Å)	0.015	0.0047
R.m.s. bond angles (°)	1.557	1.143

[a] R_{sym} is the unweighted R value on I between symmetry mates.

[b] $R_{work} = \Sigma_{hkl}[[F_{obs}(hkl)]-[F_{calc}(hkl)]]/\Sigma_{hkl}[F_{obs}(hkl)]$

[c] r_{free} is the crossvalidation R factor for 5% of reflections against which the model was not refined

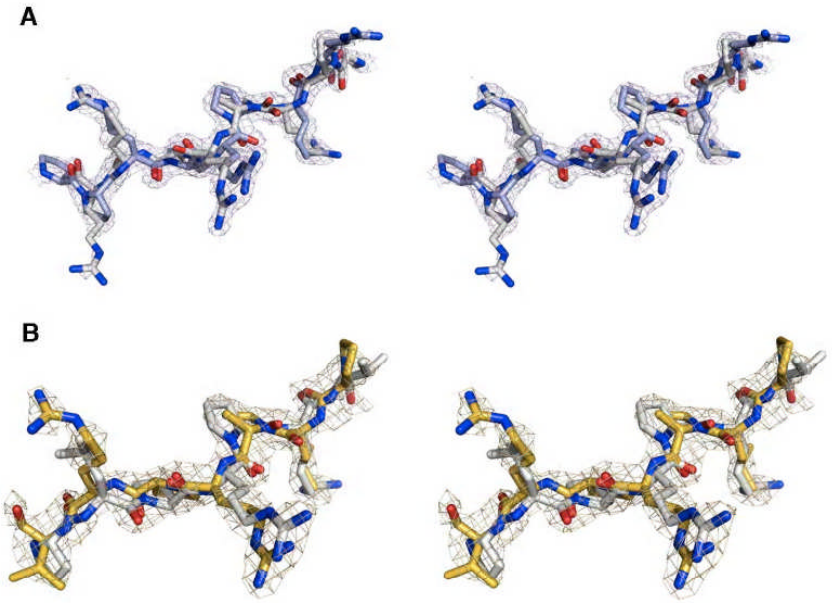
Supplemental Experimental Procedures

Purified CMSA protein (75 mg/ml final concentration in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM EDTA) crystallized at 22 ºC by mixing 1 µl of protein solution with 1 µl of reservoir solution (0.1M Sodium Citrate pH 6.1, 25%PEG4000, 20% Isopropanol). Crystals were flash cooled directly to 130 K. Dataset was collected at 130 K and processed, solved and refined as described in Experimental Procedures

In CMSA:Cbl-b P908A, prior to crystallization, 3 µmol of Cbl-b P908A peptide (PARPPKARPRR, Genosys) was added to 1.2 µmol of purified CMSA protein (70 mg/ml final concentration in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM DTT and 0.5 mM EDTA). CMSA:Cbl-b P908A crystallized at 20 ºC by mixing 1 µl of protein solution with 1 µl of reservoir solution (20% PEG3000, 0.1 M acetate pH 5.5). Crystals were transferred to stabilization buffer (10 % (v/v) ethylene glycol, 18 % PEG3000, 0.09 M acetate pH 5.5) and flash cooled to 110 ºK in the case of CMSA:Cbl-b and 298 ºK for CMSA:Cbl-b P908A. Dataset was collected at 298 K and processed, solved and refined as described in Experimental Procedures.

CMS N-terminal SH3 domain in complex with Cbl-b and CD2

Supplemental Figure



Supplemental Figure. Stereo view of symmetrically related peptides from the CMSA:Cbl-b (**A**) and CMSA:CD2 (**B**) heterotrimeric complexes show the double orientation of the peptides. In both representations one of the symmetrically related peptides is colored in grey and the other one is colored in light blue in the Cbl-b peptide and in yellow in the case of the CD2 peptide. SigmaA weighted electron density map around them is contoured at 1.0 σ.